Persistence Analysis of Hedge Fund Returns *

Serge Patrick Amvella, Iwan Meier, Nicolas Papageorgiou
HEC Montréal

November 20, 2009

Abstract

We use a Markov chain model to evaluate pure persistence in hedge fund returns. We study two forms of pure persistence: absolute persistence (positive/negative returns) and persistence with respect to the high water mark (accounting for the amplitude of draw-downs). In the first case, we find that hedge funds in general exhibit persistence of positive returns, but no persistence of negative returns. In contrast, the results using the high water mark criterion show the presence of both positive and negative persistence. In order to account for the presence of serial correlation, we use a new approach based on the method of moments and on the model of Getzmansky et al. (2004). Our findings suggest that the smoothing contributes to an increase in absolute persistence. These results also suggest that hedge fund managers exhibit a relatively high probability of delivering positive returns, but a much weaker probability of increasing their high water mark, a consequence of the non-normal distribution of their returns. Our approach also overcomes the issue of a "strategic" discontinuity in the return distribution around zero that Bollen and Pool (2009) identify and attribute to the fact that managers will adjust reported returns to minimize the chance of small negative returns in order to promote the appearance of pure persistence.

*Corresponding author: Serge Patrick Amvella, Finance Department, HEC Montréal, 3000 Cote Sainte-Catherine, Montreal,QC, H3T 2A7, Canada. All the authors are at HEC Montréal and can be reached at firstname.lastname@hec.ca. The authors are grateful to seminar participants at the 2009 EFA annual meetings in Washington D.C. and the 2009 EFMA meetings in Milan for many helpful comments. They would also like to thank Bruno Rmillard for his helpful suggestions. The authors gratefully acknowledge financial support by the Centre de Recherche en E-Finance (CREF), lInstitut de Finance Mathmatique de Montral (IFM2) and Desjardins Global Asset Management.

1 Introduction

The last few years have provided a challenging environment for hedge fund managers. As the number of hedge funds approaches the 10,000 milestone and assets under management have already surpassed the two trillion dollar mark, it is only natural that investors have become increasingly skeptical of the ability of the hedge fund industry to continue offering significant value. The absolute returns that have long been advertised by hedge fund managers have been increasingly hard to come by over the last few years, and it is estimated that approximately 80% of hedge funds were in the red during 2008. The increased market volatility, the subprime debacle and the ensuing credit crunch have recently added to an already difficult investment environment. However, given the exorbitant fee structure of these funds, investors have come to expect strong performance regardless of market conditions. The performance of these funds has been scrutinized by both practitioners and academics, and hedge fund managers are increasingly suspected of selling beta returns (returns linked to readily available market risk premia) as opposed to alpha (absolute) returns. Given the changing nature of the hedge fund universe, it is vital to identify those managers who can systematically provide positive returns, also referred to as pure persistence.

In the area of persistence evaluation, a distinction must be made between relative persistence and pure persistence. In evaluating relative persistence, funds of the same strategy are classified as winners or losers depending on their performance relative to the median return over a given period. Evidence of persistence is found when winners and/or losers maintain their classification for two subsequent periods. Most of the studies in hedge fund literature address the question of persistence in terms of relative persistence and adopt many of the tests employed in mutual fund literature where this notion has been widely explored. Relative persistence studies provide a general picture of whether past performance is a reliable indicator of future performance within a peer-group comparison framework. It doesn't isolate a specific fund and analyze its performance over time; this is achieved by investigating pure persistence. Pure persistence aims to identify funds that systematically generate positive returns. Although the study of pure persistence may be informative in the mutual fund context, it doesn't have the same relevance as relative persistence in that mutual fund managers are index trackers and are evaluated relative to their benchmark. Losses incurred by mutual fund managers are not necessarily classified as bad as long as the managers outperform their benchmark; the fact that managers are not evaluated relative to an exogenous threshold explains why there is no significant literature on pure persistence in mutual fund performance. Nonetheless, even if the studies on persistence analysis in hedge fund performance followed the same trend, it is important to note that the managers are not evaluated in the same manner. Hedge funds are absolute returns strategies and investors expect absolute returns (good returns) regardless of the market's direction. The high incentive fees charged by hedge fund managers (which average 20%) are then supposed to justify this privilege and the latter are not evaluated relative to a benchmark, but on their ability to deliver absolute returns. The fact that recent studies (among which Hasanhodzic and Lo (2007)) show that a larger proportion of hedge funds are exposed to beta driven returns calls into question the high level of incentive fees charged to investors. In the case of hedge funds, the analysis of pure persistence provides a more appropriate measure than relative persistence analysis, and allows us to identify managers exhibiting superior skills in terms of absolute performance; and in the current context of financial crisis where investors are increasingly aware of the fact that finding a manager able to deliver absolute returns is a challenge, pure persistence analysis becomes more relevant than ever.

As mentioned above, the majority of studies investigate relative persistence in hedge fund returns. Brown, Goetzmann and Ibbotson (1999), Agarwal and Naik (2000) and Liang (2000) use parametric tests (cross-sectional regressions) and non-parametric tests (Cross Product Ratio, Chi-square test, Kolomogorov-Smirnov test) to investigate the presence of relative persistence in hedge fund returns. They find no evidence of relative persistence at annual horizons even if Agarwal and Naik (2000) find that hedge fund returns persist in the short term. More recently, Kosowski, Naik and Teo (2007) use a Bayesian approach to improve the accuracy of alpha estimates. They find evidence of long term relative persistence and argue that one reason why the previous studies did not find the same results is that they relied on relatively imprecise performance measures. As for pure persistence, De Souza and Gokcan (2004) use the Hurst exponent combined with a D-statistic to study a relatively small sample of funds. They find that the funds exhibiting the strongest persistence of positive returns during the in-sample period (36 months) showed a better risk-adjusted profile in the out-of-sample period. However, the accuracy of the results remains a problem in their evaluation because one of the disadvantages with the Hurst exponent is that it requires a large sample to obtain significant results.

In this paper, we address the performance of hedge funds in terms of pure persistence. The contribution of our study is threefold. Firstly, we evaluate pure persistence in hedge funds with a new approach using a Markov chain model. Persistence is then evaluated in terms of transition probabilities. These probabilities have the advantage of not assuming an a-priori distribution of returns and are easily interpretable. Moreover, we define two types of persistence for our analysis: absolute persistence (positive/negative returns) and persistence with respect to the high water mark. It is well known that several hedge fund strategies, in particular arbitrage strategies, tend to generate positive returns of small amplitude; but when they face losses, the latter are often of larger amplitude. The analysis of absolute persistence does not capture this aspect because it does not take into account the amplitude of positive or negative returns and focuses only on the sign of returns. It follows that two managers exhibiting the same sequence of positive and negative returns over a given period would obtain the same evaluation in terms of absolute performance, regardless of the fact that one may have incurred substantially greater losses. One way to address this issue is to take into account the size of returns and to evaluate persistence with respect to the high water mark. The high water mark represents the greatest value reached by an investment during a period. A manager who tends to generate small, positive returns but faces large losses during the investment period will have trouble surpassing his high water mark. It could take considerable time for certain managers to reach their high water mark after a significant drawdown. The analysis of persistence with respect to the high water mark will then consist of assessing the ability to sustainably increase the high water mark.

Secondly, we develop a method to test the significance of persistence estimates according to the length of the sample. This helps to avoid the problem one may face when using the Hurst exponent in small samples. For this purpose, we use a one-tailed t-test which makes it possible to see whether a transition probability is statistically superior to 0.5.

Finally, we evaluate persistence before and after taking into account the serial correlation in hedge fund returns. Several studies (Asness, Krail and Liew (2001), Brooks and Kat (2002), Okunev and White (2003), Getmansky, Lo and Makarov (2004)) identify the presence of significant serial correlation in hedge fund returns, which basically leads to an underestimation of their real risk. Getmansky et al. (2004) argue that the most likely source of serial correlation in hedge fund returns is the smoothing of returns due to illiquidity and to the managers' personal motivation to optimize their performance over several periods. Illiquidity because many hedge strategies invest in illiquid assets such as non-quoted assets in private equity, some emerging market stocks and bonds, real estate and infrastructure, etc. In the event managers smooth reported returns, the disclosed volatility will be smaller than the realized volatility and hence, would upwardly bias the measure of pure persistence. Getmansky et al. (2004) propose an econometric model based on an MA(2) approach to unsmooth returns. Their model assumes that the observed return is a weighted average of "true" returns. Okunev and White (2003) use a method developed by Geltner (1993) in order to obtain a new corrected series. In this study, we use a model based on the method of moments to unsmooth returns. The advantage of our model is that it allows us to determine if it is possible to obtain satisfactory solutions (positive weights) when one tries to unsmooth returns. Indeed, hedge fund returns don't have the same order of serial correlation, and imposing an order of serial correlation for all funds as in Getmansky et al. (2004) could lead to unsatisfactory results. In their paper, they obtain negative weights for some funds whereas theoretically, and according to the assumption of their model, all weights should be positive. They argue that this can be attributed to a mis-specification of the model and that a different unsmoothing model may be more appropriate. In addition, contrary to the model of Getmansky et al. (2004), our model doesn't assume normality for the estimation of weighting coefficients.

Recently, Bollen and Pool (2009) raise an important issue regarding the reporting of hedge fund returns. Specifically, they identify the presence of a discontinuity in the distribution of returns around zero, implying that managers will adjust reported returns to minimize the chance of small negative returns in order to promote the appearance of pure persistence. The test that they propose is a t-test that measures whether the frequency of returns just below and above zero are different than those expected given the smoothed kernel estimate of the underlying distribution. The test is similar to the one used by Burgstahler and Dichev (1997) who document a similar discontinuity in the distribution of corporate earnings. Although our raw data show a sharp discontinuity in the distribution of reported returns at zero, the discontinuity disappears for the unsmoothed returns, indicating that our unsmoothing procedure eliminates the concern regarding discontinuity in the distribution of hedge fund returns.

Our study reports interesting results. First, we find that even if the smoothing of returns can contribute to increase the absolute persistence, it is not necessary to unsmooth the returns of all funds. Strategies that invest in liquid securities generally exhibit a lower level of serial correlation. Our results show that imposing a MA(2) model to unsmooth their returns leads to an underestimation of the volatility of true returns. Second, the results based on our sample data suggest that hedge fund managers exhibit a relatively high probability of positive returns. However, even if negative returns don't persist, the managers exhibit some difficulties in increasing the investor's wealth in a sustainable way because periods of positive returns are sometimes interrupted by large drawdowns. When we account for this asymmetry in returns, through the persistence analysis with respect to the high water mark, we find a much weaker probability of increasing the high water mark in comparison with the probably of delivering positive returns. This interesting finding suggests that the persistence with respect to the high water mark is most effective than the absolute persistence analysis because it accounts for the particular profile of hedge fund returns and indicates the manager's ability to sustainably increase the investor's wealth.

The rest of the paper is organized as follows. Section 2 describes the methodology used to test the significance of the transition probabilities and in section 3, we present the methodology used to unsmooth returns. Section 4 presents the data and section 5 shows the results of the analysis. We conclude the study in section 6.

2 Methodology to measure pure persistence

Contrary to De Souza and Gokcan (2004), pure persistence will firstly be evaluated herein in terms of the probability of positive or negative returns over two periods. There are many advantages of using probabilities in the performance evaluation. They make no assumptions as to the distribution of returns and are more easily interpretable for an investor than the combined analysis of the Hurst exponent and the D-statistic. Moreover, probabilities allow for an approximation of the odds that a fund obtains desirable returns, which is not the case for other measures such as the mean of returns. The mean may provide the average performance of a manager over a period, but it doesn't indicate how the manager performs on a regular basis. For example, an average of 2% indicates that on aggregate, the manager's performance is above zero, but it does not indicate at which frequency he obtained positive returns or what his odds are of providing positive returns. For instance, a fund could exhibit the following returns: -2%, -1%, 15%, -1.2% -0.8%. This gives a positive mean return of 2%, but the fund experiences negative returns 4 months out of 5, (with a probability of 80%.)

Another advantage of using probabilities in the relation between past and future returns is that contrary to serial correlation, which is only relevant for elliptical distributions and measures the linear dependence between the returns, probabilities apply to other distributions and can measure dependence that may be non-linear; and we know from available literature that hedge fund returns are often non-Gaussian due to the use of derivatives and dynamic trading strategies (Fung and Hsieh (1997), Agarwal and Naik (2004), etc.).

The evaluation of persistence is done through a Markov chain model. Persistence is then measured in terms of transition probabilities. A Markov chain is a stochastic process where the prediction of the future depends on the present and is independent of the past. The set of possible values that the random variable can take is referred to as the state space and the Markovian property is defined as follows:

$$\Pr[X_{t+1} = j | X_0 = i_0, ..., X_{t-1} = i_{t-1}, X_t = i] = \Pr[X_{t+1} = j | X_t = i] \tag{1}$$

where t represents the time for the states $i_0, \ldots, i_{t-1}, i, j$. We will use a two-state Markov chain to evaluate persistence. Let R_t, denote the return of the fund at time t and I_t a dichotomous variable that follows the process:

$$\begin{aligned} I_t &= 1 \;\; if \;\; R_t > 0 \\ I_t &= 0 \;\; if \;\; R_t \leq 0 \end{aligned} \tag{2}$$

The series derived from this transformation follows a two-state Markov chain and identifies strictly positive returns as 1 and negative or null returns as 0. The corresponding transition matrix is:

$$M = \begin{bmatrix} p_{11} & p_{10} \\ p_{01} & p_{00} \end{bmatrix}$$

with

$$\begin{aligned} p_{11} &= \Pr[I_{t+1} = 1 | I_t = 1] \\ p_{10} &= \Pr[I_{t+1} = 0 | I_t = 1] \\ p_{01} &= \Pr[I_{t+1} = 1 | I_t = 0] \\ p_{00} &= \Pr[I_{t+1} = 0 | I_t = 0] \end{aligned}$$

The elements in the diagonal of the transition matrix (p_{11} and p_{00}) identify the presence of positive and negative persistence of returns. p_{01} and p_{10} indicate the probabilities of obtaining a gain after a loss, and vice versa. The transition probabilities are calculated to maximize the following likelihood function:

$$L(S_T,\ p_{i,}\ \pi) = \log \pi + \sum_{ij=00}^{11} N_{ij} \log p_{ij} \ + M_{ij} \log(1 - p_{ij}) \tag{3}$$

where S_T is the set of realized I_t, and π the probability of the initial state. The latter can take the following values:

- If the initial state $I_1 = 1$

$$\pi = \pi_1 = \frac{1 - p_{00}}{2 - p_{11} - p_{00}} \tag{4}$$

- If the initial state $I_1 = 0$

$$\pi = \pi_0 = \frac{1 - p_{11}}{2 - p_{11} - p_{00}} \tag{5}$$

N_{ij} *and* M_{ij} are the occurrences associated with the various transitions. It is important to notice that π is a function of the transition probabilities[1].

[1] For more information, refer to *Time Series Analysis*, J. D. Hamilton, Princeton University, 1994.

In this context of persistence analysis of hedge fund returns with limited historical data, it is important to ensure the significance of the transition probabilities. For this purpose, we developed an approach to test whether or not persistence estimators are statistically significant. To our knowledge, the existing tests in the literature for Markov chains consist mostly of independence or random walk tests and are generally based on likelihood ratio tests or χ^2- tests[2]. For example, we know that $p_{11} > 0.5$ indicates positive persistence and $p_{00} > 0.5$ indicates negative persistence. Therefore, testing for positive persistence is equivalent to performing the following unilateral test:

$$\begin{aligned} &H_0\text{: } p_{11} \leq 0.5 \\ &H_1\text{: } p_{11} > 0.5 \end{aligned}$$

The corresponding t-statistic is:

$$t = \frac{\hat{p}_{11} - 0.5}{\hat{\sigma}_{p_{11}}} \sim t_c(n-1) \tag{6}$$

Hence, we require the volatility estimate $\hat{\sigma}_{p_{11}}$. To this end, we firstly estimate the asymptotic value of $Var\left[\sqrt{n}\left(\widehat{p}_{11} - p_{11}\right)\right]$ where p_{11} is the asymptotic value of the transition probability. This is achieved via the Delta method described below. We know that $\hat{p}_{11}$ can also be expressed as follows:

$$\hat{p}_{11} = \frac{\hat{P}_{11}}{\hat{P}_{11} + \hat{P}_{10}} \tag{7}$$

where $\hat{P}_{11} = \Pr(I_t = 1;\, I_{t+1} = 1)$ and $\hat{P}_{10} = \Pr(I_t = 1;\, I_{t+1} = 0)$ are jointed probabilities. Thus, $\hat{p}_{11}$ is a function of $\hat{P}_{11}$ and $\hat{P}_{10}$ and we can write:

$$\hat{p}_{11} = f(\hat{P}_{11}, \hat{P}_{10})$$

By the Delta method, and with some assumptions, we can show that[3]:

$$Var\left[\sqrt{n}\left(\widehat{p}_{11} - p_{11}\right)\right] = Var\left[\sqrt{n}\left(\widehat{P}_{11}\right)\right] + Var\left[\sqrt{n}\left(\widehat{P}_{10}\right)\right] - 2Cov\left[\sqrt{n}\left(\widehat{P}_{11}\right), \sqrt{n}\left(\widehat{P}_{10}\right)\right] \tag{8}$$

In the appendix, we show that when $n \longrightarrow \infty$:

$$\begin{aligned} Var\left(\sqrt{n}\widehat{P}_{11}\right) &\longrightarrow \frac{5}{16} \\ Var\left(\sqrt{n}\widehat{P}_{10}\right) &\longrightarrow \frac{1}{16} \\ Cov\left(\sqrt{n}\widehat{P}_{11}, \sqrt{n}\widehat{P}_{10}\right) &\longrightarrow -\frac{1}{16} \end{aligned}$$

[2]The reader can refer to the work of P.G. Hoel, L. A. Goodman, C. K. Tsao and other authors. Some tests for Markov chains can be found in the following papers: P.G. Hoel (1954) "A test for Markoff Chains" , *Biometrika*, **41** pp. 430-433; Goodman, L. A. (1958) "Simplified Runs Tests and Likelihood Ratio Tests for Markov Chains", *Biometrika*. **51** pp. 89-100 ; Tsao, C. K. (1968) " Admissibility and Distribution of Some Probabilistic Functions of Discrete Finite State Markov Chains", *Ann. Math. Statist.* **39** pp. 1646-1653.

[3]The demonstration can be found in appendix A.

This gives

$$Var\left[\sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)\right]=\frac{1}{2}$$

From this result and the central limit theorem, the following can be obtained[4]:

$$\sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)\rightarrow N(\ 0,\ \frac{1}{2}\ )$$

Therefore

$$\hat{\sigma}_{p_{11}}=\ \frac{1}{\sqrt{2n}} \tag{9}$$

We follow the same procedure for p_{00} (in appendix) and the results show that

$$\sqrt{n}\left(\widehat{p}_{00}-p_{00}\right)\rightarrow N(\ 0,\ \frac{1}{2}\ )$$

and

$$\hat{\sigma}_{p_{00}}=\frac{1}{\sqrt{2n}} \tag{10}$$

3 Methodology to unsmooth returns

In this study, we estimate persistence for the smoothed and unsmoothed returns of each fund. This enables us to verify whether the smoothing of returns has an effect on persistence and if so, which strategies are the most affected. Getmansky, Lo and Makarov (2004) (henceforth GLM) propose a model using maximum likelihood estimation to obtain the "unsmoothed" time series of returns. The model of GLM assumes that the observed return in period t (R_t^o) is a weighted average of the "true" returns (R^c) over the most recent k +1 periods, including the current period:

$$R_t^o=\theta_0 R_t^c+\theta_1 R_{t-1}^c...+\ \theta_k R_{t-k}^c \tag{11}$$

$$\theta_j\ \epsilon\ [0,1]\,,\quad j=0,...,k \tag{12}$$

$$1=\theta_0+\theta_1...+\ \theta_k \tag{13}$$

The θs can be estimated using the maximum likelihood approach. The smoothing level (or smoothing index) is equal to the sum of the squared θ_j:

$$\xi=\sum_j^k\theta_j^2 \tag{14}$$

By construction $0\leq\ \xi\ \leq 1$. A small value of ξ implies a high smoothing level, ξ =1 indicates no smoothing. After estimating the θs, the "true" returns (unsmoothed) are obtained by inverting the equation in this way:

[4] We performed a bootstrap with a large sample of data and the variance converges towards 1/2.

$$R_t^c = \frac{R_t^o - \hat{\theta}_1 R_{t-1}^c ... - \ \hat{\theta}_k R_{t-k}^c}{\hat{\theta}_0} \tag{15}$$

The unsmoothed and the observed returns have the same mean, but not the same variance. The variance of the unsmoothed returns is higher than that of the observed returns ($\sigma_c^2 \geq \sigma_o^2$) and the relation between both variances is as follows: $\sigma_o^2 = \xi \sigma_c^2$.

To estimate the θs, GLM first centered the observed returns to come up with a new time series:

$$X_t = R_t^o - \mu \tag{16}$$

Given the process described before the equation becomes:

$$X_t = R_t^o - \mu = \theta_0(R_t^c - \mu) + \theta_1(R_{t-1}^c - \mu)... + \ \theta_k(R_{t-k}^c - \mu) + (\theta_0 + \theta_1 ... + \ \theta_k)\mu - \mu$$

Setting $R_t^c - \mu = \eta_t, \quad R_{t-1}^c - \mu = \eta_{t-1}, \ ... \quad R_{t-k}^c - \mu = \eta_{t-k},$ we get :

$$X_t = \theta_0 \eta_t + \theta_1 \eta_{t-1} ... + \ \theta_k \eta_{t-k} \tag{17}$$

$$1 = \theta_0 + \theta_1 ... + \ \theta_k \tag{18}$$

$$\eta_t \sim N(0, \sigma_\eta^2) \tag{19}$$

where the last assumption is added for purposes of estimation of the *MA(k)* process.

In their model, GLM estimate the θs for 909 hedge funds with a MA(2) assuming a serial correlation of lag 2 for hedge fund returns. This method is very attractive but nevertheless raises some problems. On the one hand, it is based on the assumption that demeaned returns (η_t) follow a normal distribution and the authors mention that although the maximum likelihood estimation has some attractive properties it is only consistent and asymptotically efficient under certain regularity conditions. Therefore, it may not perform well in small samples or when the underlying distribution of true returns is not normal. Moreover, GLM mention that even if the normality condition is satisfied and a sufficient sample size is available, the smoothing model simply may not apply to certain funds. If the numerical optimization does not converge it could be due to the fact that the model is mis-specified, due to either non-normality or an inappropriate specification of the model. Another check is to verify whether or not the estimated smoothing coefficients are all positive in sign. Estimated coefficients that are negative and significant may be a sign that the constraint of positivity (of weights) is violated, which suggests that a somewhat different smoothing model may apply. In their study which imposes an MA(2) specification, they obtain negative weights (negative values for θ_1 and θ_2) for some funds. It is important to note that not all funds have the same level of serial correlation and therefore, imposing the same level of serial correlation for all funds could lead to the estimation of mis-specified parameters θ_j and this could have undesirable effects on the distribution of unsmoothed returns. For example, when a parameter θ_j is negative, the fact that the weights must sum to 1 implies that at least one of them should be greater than 1. In this case, we would have a smoothing level $\xi > 1$ and the variance of unsmoothed returns would be lower than the variance of the observed returns, which would underestimate the true risk of the fund. This suggests that it is very important to specify the appropriate model for each fund. For example, funds investing in liquid securities will probably have serially uncorrelated returns and imposing the unsmoothing of their returns could lead to mis-specified θs. This is why it is important to firstly check the level of the serial correlation of returns.

In this study, we propose a model based on the method of moments to estimate the θs. Our model has the advantage of identifying when it is possible to obtain a satisfactory solution for θs. In addition, our model doesn't assume normality; this is a relevant point given that many studies documented the non-normality of hedge fund returns. Let us reconsider the model of GLM (2004):

$$X_t = \theta_0 \eta_t + \theta_1 \eta_{t-1} ... + \ \theta_k \eta_{t-k} \tag{20}$$

$$1 = \theta_0 + \theta_1 ... + \ \theta_k \tag{21}$$

$$\eta_t \sim D(0, \sigma_\eta^2) \tag{22}$$

In this case, the demeaned η_t follows a distribution D which is not necessarily normal. We only suppose that the unobserved returns are independent and have a constant volatility to estimate. Suppose the observed returns are serially correlated up to lag k. By using the method of moments, it implies:

$$\begin{aligned} E\left[X_t^2\right] &= E\left[(\theta_0\eta_t + \theta_1\eta_{t-1} ... + \ \theta_k\eta_{t-k}).(\theta_0\eta_t + \theta_1\eta_{t-1} ... + \ \theta_k\eta_{t-k})\right] \\ &= \theta_0^2\sigma_\eta^2 + \theta_1^2\sigma_\eta^2 + ... + \ \theta_k^2\sigma_\eta^2 \\ &= (\theta_0^2 + \theta_1^2 + ... + \ \theta_k^2)\sigma_\eta^2 \end{aligned}$$

$$\begin{aligned} E\left[X_t.X_{t-1}\right] &= E\left[(\theta_0\eta_t + \theta_1\eta_{t-1} ... + \ \theta_k\eta_{t-k}).(\theta_0\eta_{t-1} + \theta_1\eta_{t-2} ... + \ \theta_k\eta_{t-k-1})\right] \\ &= \theta_0\theta_1\sigma_\eta^2 + \theta_1\theta_2\sigma_\eta^2 + ... + \ \theta_{k-1}\theta_k\sigma_\eta^2 \\ &= (\theta_0\theta_1 + \theta_1\theta_2 + ... + \ \theta_{k-1}\theta_k)\sigma_\eta^2 \end{aligned}$$

$$\begin{aligned} E\left[X_t.X_{t-2}\right] &= E\left[(\theta_0\eta_t + \theta_1\eta_{t-1} ... + \ \theta_k\eta_{t-k}).(\theta_0\eta_{t-2} + \theta_1\eta_{t-3} ... + \ \theta_k\eta_{t-k-2})\right] \\ &= \theta_0\theta_2\sigma_\eta^2 + \theta_1\theta_3\sigma_\eta^2 + ... + \ \theta_{k-2}\theta_k\sigma_\eta^2 \\ &= (\theta_0\theta_2 + \theta_1\theta_3 + ... + \ \theta_{k-2}\theta_k)\sigma_\eta^2 \end{aligned}$$

...

$$\begin{aligned} E\left[X_t.X_{t-k}\right] &= E\left[(\theta_0\eta_t + \theta_1\eta_{t-1} ... + \ \theta_k\eta_{t-k}).(\theta_0\eta_{t-k} + \theta_1\eta_{t-k-1} ... + \ \theta_k\eta_{t-2k})\right] \\ &= \theta_0\theta_k\sigma_\eta^2 \end{aligned}$$

Thus, we have k *+1* moment conditions, and we want to estimate k *+2* parameters. We also have one more condition, which is $\sum_j^k \theta_j = 1$. This leads to a system of k *+2* equations with k *+2* unknown parameters:

$$\left\{\begin{array}{r l} E\left[X_t^2\right] &= (\theta_0^2+\theta_1^2+...+\theta_k^2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] &= (\theta_0\theta_1+\theta_1\theta_2+...+\theta_{k-1}\theta_k)\sigma_\eta^2 \\ E\left[X_t.X_{t-2}\right] &= (\theta_0\theta_2+\theta_1\theta_3+...+\theta_{k-2}\theta_k)\sigma_\eta^2 \\ & ... \\ E\left[X_t.X_{t-k}\right] &= \theta_0\theta_k\sigma_\eta^2 \\ 1 &= \theta_0+\theta_1...+\theta_k \end{array}\right. \tag{23}$$

We are then able to estimate the parameters. One simple way to do this is to firstly estimate the order k of serial correlation of the observed returns. In the GLM model, they assume that all the funds have returns serially correlated up to lag 2, which is not necessarily true. For example, Managed futures funds, for the most part, have serially uncorrelated returns because they generally invest in liquid securities; imposing a level of serial correlation could lead to mis-specified parameters. Our approach is to firstly measure the level of serial correlation and then estimate the corresponding parameters θ_j and σ_η^2. We will limit the development to lag 2. Depending on the level of serial correlation found, we have three main cases:

a) First case: k = 0

If the first- and the second-order serial correlation are not statistically significant, it is not necessary to unsmooth the returns and we keep them as they are.

b) Second case: k = 1

If the first-order serial correlation is statistically significant but not the second one, we have 3 parameters to estimate θ_0, θ_1 and σ_η^2 from the following system of equations:

$$\left\{\begin{array}{l} E\left[X_t^2\right] = (\theta_0^2+\theta_1^2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = \theta_0\theta_1\sigma_\eta^2 \\ 1 = \theta_0+\theta_1 \end{array}\right. \tag{24}$$

The resolution of this system of equations gives the following results[5]:

$$\sigma_\eta^2 = E\left[X_t^2\right]+2.E\left[X_t.X_{t-1}\right] \tag{25}$$

$$\theta_0 = \frac{1}{2}+\frac{\sqrt{1-4\gamma_1}}{2} \tag{26}$$

$$\theta_1 = \frac{1}{2}-\frac{\sqrt{1-4\gamma_1}}{2} \tag{27}$$

with

$$\gamma_1=\frac{E\left[X_t.X_{t-1}\right]}{\sigma_\eta^2} \tag{28}$$

[5]The developments are presented in appendix C.

Then, the system's solutions exist if and only if $\gamma_1 \leq \frac{1}{4}$ and to obtain satisfactory solutions, $(\theta_1 \geq 0)$, γ_1 should lead in this interval:

$$0 \leq \gamma_1 \leq \frac{1}{4} \tag{29}$$

The first-order serial correlation should not be too high, nor should it be negative because if $\gamma_1 < 0$ i.e. if $Cov(X_t, X_{t-1}) < 0$, we will have $\theta_1 < 0$. In other words, if the first-order serial correlation is negative, not all weights will be positive and the unsmoothing will be incongruous because ξ will be higher than 1 and σ_c^2 will be lower than σ_o^2. Note that σ_η^2 and γ_1 can empirically be estimated from the sample equivalent of $E\left[X_t^2\right]$ and $E\left[X_t X_{t-1}\right]$.

c*) Third case: k = 2*

If the first- and the second-order serial correlation are both statistically significant we have 4 parameters to estimate θ_0, θ_1, θ_2 and σ_η^2 from the following system of equations:

$$\left\{\begin{array}{l} E\left[X_t^2\right] = (\theta_0^2+\theta_1^2+\theta_2^2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = (\theta_0\theta_1+\theta_1\theta_2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = \theta_0\theta_2\sigma_\eta^2 \\ 1 = \theta_0+\theta_1+\theta_2 \end{array}\right. \tag{30}$$

The resolution of this system of equations gives the following results:

$$\sigma_\eta^2 = E\left[X_t^2\right] + 2.E\left[X_t.X_{t-1}\right] + 2.E\left[X_t.X_{t-2}\right] \tag{31}$$

$$\theta_1 = \frac{1}{2} - \frac{\sqrt{1-4\delta_1}}{2} \tag{32}$$

$$\theta_0 = \frac{(1-\theta_1)}{2} + \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2} \tag{33}$$

$$\theta_2 = \frac{(1-\theta_1)}{2} - \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2} \tag{34}$$

with

$$\delta_1 = \frac{E\left[X_t.X_{t-1}\right]}{\sigma_\eta^2} \tag{35}$$

$$\delta_2 = \frac{E\left[X_t.X_{t-2}\right]}{\sigma_\eta^2} \tag{36}$$

Then, the system's solutions exist if and only if $\delta_1 \leq \frac{1}{4}$ and $\delta_2 \leq \frac{(1-\theta_1)^2}{4}$. To have satisfactory solutions, δ_1 and δ_2 should lead in these intervals:

$$0 \leq \delta_1 \leq \frac{1}{4} \tag{37}$$

$$0 \leq \delta_2 \leq \frac{(1-\theta_1)^2}{4} \tag{38}$$

The first- and the second-order serial correlation should not be too high, nor should they be negative because if $\delta_1 < 0$ (i.e. if $Cov(X_t, X_{t-1}) < 0$) and/or if $\delta_2 < 0$ ($Cov(X_t, X_{t-2}) < 0$), we will have $\theta_1 < 0$ and /or $\theta_2 < 0$ and there is a possibility that θ_0 may be greater than 1, and ξ then also greater than 1. In other words, if one or both of the serial correlations is negative, not all weights will be positive and the unsmoothing will be incongruous because ξ will be greater than 1, and σ_c^2 will be less than σ_o^2. Note that σ_η^2, δ_1 and δ_2 can empirically be estimated from the sample equivalent of $E\left[X_t^2\right]$, $E\left[X_t X_{t-1}\right]$ and $E\left[X_t X_{t-2}\right]$.

d) Decision process

Before evaluating pure persistence for each fund, we calculate the first- and the second-order serial correlation of returns and the decision process is as follows:

(*) If neither is statistically significant, we keep the observed returns.

(**) If only the first-order serial correlation is significant (k=1), we estimate σ_η^2 and γ_1, and:

- If $0 \leq \gamma_1 \leq \frac{1}{4}$, we estimate θ_0 , θ_1 and the unsmoothed returns as follows:

$$R_t^c = \frac{R_t^o - \hat{\theta}_1 R_{t-1}^c}{\hat{\theta}_0} \tag{39}$$

Note that if $k = 1$, the estimation of the unsmoothed returns is based on the assumption that the first return is an unsmoothed return.

- $\gamma_1 < 0$ implies that it is not possible to obtain satisfactory solutions and we exclude the fund from our sample.

- $\gamma_1 > \frac{1}{4}$ implies that the first-order serial correlation is too high, and we therefore estimate the model as if $k = 2$ to see whether we can obtain a solution. If not, we exclude the fund from our sample.

(***) If both the first- and the second-order serial correlations are statistically significant, we estimate σ_η^2, δ_1, δ_2 and θ_1, and verify that $0 \leq \delta_1 \leq \frac{1}{4}$ and $0 \leq \delta_2 \leq \frac{(1-\theta_1)^2}{4}$. In this case, we estimate θ_0 , θ_2 and the unsmoothed returns as follows:

$$R_t^c = \frac{R_t^o - \hat{\theta}_1 R_{t-1}^c - \hat{\theta}_2 R_{t-2}^c}{\hat{\theta}_0} \tag{40}$$

If δ_1 and δ_2 are not comprised within these intervals, we exclude the fund from our sample because we can not obtain satisfactory solutions, or we can not obtain a solution at all.

3.1 Robustness check for discontinuity

A natural question that arises is to know whether our unsmoothing procedure clears the concern of discontinuity in hedge fund returns reported by Bollen and Pool (2009). They examine the histogram of the pooled distribution of reported hedge fund returns and find that it exhibits a discontinuity at zero, i.e. returns just below zero are under-represented and returns just above zero are over-represented, suggesting that some managers distort returns when possible in order to avoid reporting losses. They also find that this phenomenon seems to be more pronounced in hedge funds styles that focus in illiquid securities. Getzmansky et al. (2004) argue that illiquidity and the managers' personal motivation to optimize their performance are the main sources of the smoothing of returns. So, funds with more smoothing can be considered as those with more illiquid assets, but can also be considered as those where there is more distortion of returns. However, even if the distortion of returns is more feasible when the manager is invested in illiquid assets, it remains difficult to distinguish the discontinuity created by purposeful smoothing and that created by innocuous smoothing.

To investigate the presence of discontinuity around zero, Bollen and Pool (2009) use a test similar to that of Burgstahler and Dichev (1997) who document a discontinuity in the distribution of corporate earnings of firms listed in the Compustat database. The t-test measures whether the height of the bins adjacent to zero are consistent with the smoothed kernel estimate of the underlying distribution. In order to evaluate whether our smoothing procedure eliminates the problem of discontinuity, we implement the same test on the pooled distribution of funds whose returns have been unsmoothed - that is funds whose reported returns exhibit a statistically significant first or second order serial correlation[6]

4 Data

Our hedge funds data comprises the monthly net-of-fee returns of 7,255 live and dead funds provided by Hedge Fund Research Inc. (HFR) and covers the period starting January 1994 and ending December 2007. However, we excluded funds with less than 36 consecutive monthly returns in order to estimate pure persistence with sufficient data. This led us to a total of 4,783 funds. Our data consists of 20 hedge fund strategies and is representative of the hedge fund universe. Table 1 exhibits the statistics of funds for different strategies and the values presented are the average values across the strategies. We can see that there is an unequal distribution of funds in various strategies. Funds of funds are the most numerous (1,748), whereas Short selling has the lowest number of funds (13). On average, all the strategies exhibit a positive mean with the highest values for Emerging market (1.81%), Sector (1.44%) and Equity non-hedge (1.37%).

Short selling, Equity non-hedge and Emerging market exhibit the highest volatility values. With regard to the third and the fourth moment of the distribution, hedge funds exhibit skewed returns and excess kurtosis. These descriptive statistics are in line with the results found in various studies documenting the non-normality of hedge fund returns (Fung and Hsieh (1997),

[6]In order to compare our results with those of Bollen and Pool (2009), we exclude Funds of funds and Managed future funds which correspond to CTAs.

Table 1: Descriptive statistics for hedge fund returns

	Mean (%)	Vol.(%)	Skewness	Kurtosis	Funds
Convertible Arb	0.67	1.57	-0.46	5.24	92
Distress Sec.	1.11	2.31	0.33	6.41	107
Emerging Mkt	1.81	5.06	0.06	5.96	196
Equity Hedge	1.09	3.55	0.19	5.18	992
Equity Mkt N.	0.63	2.13	-0.18	6.06	193
Equity Non Hedge	1.37	5.19	0.06	5.13	121
Event Driven	1.13	2.94	0.05	6.17	174
FI Arbitrage	0.54	1.68	-0.44	9.77	70
FI Convertible	0.66	3.35	0.26	4.62	21
FI Diversified	0.60	1.88	-0.55	8.93	65
FI High Yield	0.69	1.89	-1.42	13.30	50
FI Mortgage	0.79	1.94	-1.66	20.46	38
Fund of Funds	0.70	1.68	-0.37	5.13	1747
Macro	0.96	3.62	0.07	5.06	212
Market Timing	1.05	3.88	0.64	7.70	24
Managed Futures	0.97	5.03	0.29	4.55	223
Merger Arb.	0.77	1.54	0.05	7.63	43
Relative Value	0.91	1.98	-0.04	6.69	211
Sector	1.44	4.92	0.30	6.07	191
Short Selling	0.12	6.83	0.00	6.23	13
All	0.93	2.80	-0.11	5.72	4783

This table presents the monthly mean return, volatility, skewness, and kurtosis as well as the number of funds for each hedge fund strategy

Liang (2000), etc.).

It is also well documented that hedge fund data is subject to various biases such as survivorship bias or backfill bias. We construct our data set so as to limit any exposure to these biases. By using the returns of live and dead funds, we avoid the survivorship bias given that persistence is evaluated for both successful and unsuccessful funds. In order to account for the backfill bias, some studies exclude the first 12 monthly returns as some funds may report their returns before their inclusion in the database if the returns are good. To verify whether it was necessary to use the same process on our sample, we estimated, for each fund, the difference in mean with and without the first 12 months. The values obtained are presented in table 2. $\mu_{(\text{all})} - \mu_{(\text{minus 12})}$ and $\sigma_{(\text{all})} - \sigma_{(\text{minus 12})}$ are, respectively, the differences in mean and volatility between the the entire set of the funds' returns and that which excludes the first 12 months. The average differences for each strategy and the corresponding t-statistic are presented in the table.

We can see that the differences in mean are small, and even negative for some strategies (Emerging market, FOF, Market timing and Short selling), which indicates that the mean is not necessarily increased when one includes the first 12 months of data. The spreads range from a minimum of -0.046% for Short selling to a maximum of 0.091% for Equity non-hedge. The t-statistics show that the spreads are not statistically different from zero, except for Equity hedge funds. Including the first 12 months of returns does not necessarily create a backfill bias in our database and we will therefore use all available data in our study.

Table 2: Average differences in returns when first 12 months are excluded returns

	$\mu_{(\text{all})} - \mu_{(\text{minus 12})}$	$\sigma_{(\text{all})} - \sigma_{(\text{minus 12})}$	t-stat
Convertible Arb	0.06	0.16	0.39
Distress Sec.	0.04	0.22	0.20
Emerging Mkt	0.00	0.46	-0.01
Equity Hedge	0.07	0.34	0.20
Equity Mkt Neutral	0.06	0.16	0.35
Equity Non Hedge	0.09	0.38	0.24
Event Driven	0.05	0.33	0.15
FI Arbitrage	0.06	0.16	0.34
FI Convertible	0.04	0.18	0.20
FI Diversified	0.05	0.12	0.45
FI High Yield	0.05	0.17	0.29
FI Mortgage	0.09	0.14	0.65
Fund of Funds	-0.01	0.13	-0.11
Macro	0.02	0.24	0.08
Market Timing	0.00	0.36	-0.01
Managed Futures	0.02	0.35	0.06
Merger Arb.	0.04	0.13	0.30
Relative Value	0.08	0.25	0.30
Sector	0.07	0.37	0.18
Short Selling	-0.05	0.12	0.38
All	0.03	0.26	0.11

This table presents the difference in mean and volatlity when the first 12 months are excluded. The table reports the average differences in the means and volatilities, $\mu_{(\text{all})} - \mu_{(\text{minus 12})}$ and $\sigma_{(\text{all})} - \sigma_{(\text{minus 12})}$, and the corresponding t-statistic

5 Estimation results

5.1 Serial correlation of hedge fund returns

Before proceeding with the unsmoothing of returns, we firstly analyze the serial correlation of the hedge funds in our data. Table 3 presents the first- and the second-order serial correlation of the reported returns across all strategies. Columns 5 and 9 present, for each strategy, the percentage of funds exhibiting a statistically significant serial correlation of order 1 or 2.

On average, Convertible arbitrage, Distress securities, Fixed income convertible bonds, Fixed income high yield and Fixed income mortgage exhibit a higher first-order serial correla-

Table 3: First and second order serial correlation for reported returns

	First order				Second Order			
	Mean	Min	Max	% Sig.	Mean	Min	Max	% Sig.
Convertible Arb	0.38	0.07	0.86	90.2	0.12	-0.22	0.81	23.9
Distress Sec.	0.25	-0.17	0.55	65.4	0.08	-0.24	0.52	15.0
Emerging Mkt	0.12	-0.22	0.49	24.5	0.01	-0.26	0.32	3.1
Equity Hedge	0.11	-0.35	0.71	19.7	0.01	-0.52	0.47	6.7
Equity Mkt Neutral	0.06	-0.32	0.88	17.1	0.00	-0.42	0.85	6.7
Equity Non Hedge	0.10	-0.34	0.37	18.2	0.00	-0.24	0.35	3.3
Event Driven	0.20	-0.33	0.51	47.7	0.05	-0.23	0.35	8.6
FI Arbitrage	0.15	-0.43	0.76	34.3	0.03	-0.53	0.61	11.4
FI Convertible	0.20	-0.07	0.33	57.1	0.07	-0.11	0.30	19.0
FI Diversified	0.13	-0.46	0.83	24.6	-0.05	-0.35	0.80	7.7
FI High Yield	0.28	-0.05	0.50	56.0	0.01	-0.27	0.22	0.0
FI Mortgage	0.22	-0.07	0.57	44.7	0.15	-0.12	0.50	28.9
Fund of Funds	0.17	-0.57	0.66	29.4	-0.03	-0.35	0.48	4.2
Macro	0.06	-0.29	0.40	11.3	-0.03	-0.36	0.30	2.8
Market Timing	0.08	-0.15	0.39	25.0	0.06	-0.20	0.37	25.0
Managed Futures	0.03	-0.30	0.52	6.7	-0.09	-0.44	0.42	2.2
Merger Arb.	0.17	-0.26	0.49	37.2	0.08	-0.12	0.38	16.3
Relative Value	0.18	-0.40	0.84	42.7	0.03	-0.35	0.65	11.4
Sector	0.09	-0.24	0.62	15.2	-0.02	-0.36	0.48	8.4
Short Selling	0.06	-0.13	0.34	15.4	-0.06	-0.15	0.18	0.0
S&P 500	-0.006				-0.04			

This table presents the minimum, maximum, mean estimates for first and second order serial correlation for each hedge fund style. The table also shows the percentage of funds for which the estimates that are significant at the 5% level

tion. These strategies also exhibit the higher proportion of funds with a statistically significant serial correlation. And even if the second-order serial correlation is, on average, lower across all strategies, it is higher for the previously mentioned strategies, which are generally invested in illiquid securities. One can therefore expect that the unsmoothing process may apply to most of the funds in these strategies.

Also note that the serial correlation profile can vary a lot from fund to fund in each strategy and the gap between the lowest and the highest serial correlation can be very wide. For some strategies, there are certain funds whose first- or second-order serial correlation is greater than 0.80 (Convertible arbitrage, Equity market neutral, Fixed income diversified and Relative value arbitrage). This shows that if one wants accurate results when analyzing hedge funds, it is important to work on a fund-by-fund basis rather than analyzing the aggregate data of indices. Table 3 also shows that strategies involved in more liquid securities such as Macro or Managed futures are those for which the first-order serial correlation is lower. Therefore, the unsmoothing process should be less applicable to these strategies.

The last row of the table 3 shows the first- and the second-order serial correlation of S&P500 monthly returns from January 1994 to December 2007. We can see that they are very small and not statistically significant.

5.2 Results for the un-smoothed of returns

For the sake of comparison, we proceed with to un-smooth the retruns using two approaches. First, we impose a first- and second- order serial correlation on all funds (constrained model as per that of GLM) and second, we unsmooth the returns according to the level of serial correlation of each fund (unconstrained model). Table 4 presents the average values of θ_0, θ_1, θ_2 and ξ for each strategy in the constrained model. The last column presents the percentage of funds for which we can obtain possible solutions (but not necessarily satisfactory solutions). Funds for which we have no possible solution are those for which the level of first- or second-order of serial correlation is very high or the order of serial correlation is greater than 2.

Table 4: Serial Correlation: Constrained model

	θ_0	θ_1	θ_2	ξ	% of funds
Convertible Arb	0.64	0.27	0.09	0.52	96.7
Distress Sec.	0.76	0.18	0.06	0.65	96.3
Emerging Mkt	0.91	0.09	-0.01	0.91	100.0
Equity Hedge	0.94	0.08	-0.02	1.01	99.1
Equity Mkt Neutral	1.18	-0.04	-0.14	10.65	99.5
Equity Non Hedge	0.94	0.07	-0.01	0.97	100.0
Event Driven	0.81	0.15	0.04	0.73	100.0
FI Arbitrage	0.91	0.07	0.01	1.03	92.9
FI Convertible	0.79	0.16	0.05	0.69	100.0
FI Diversified	1.02	0.05	-0.07	1.36	90.8
FI High Yield	0.77	0.24	-0.01	0.70	100.0
FI Mortgage	0.75	0.14	0.11	0.65	92.1
Fund of Funds	0.90	0.15	-0.05	0.91	98.9
Macro	1.04	0.02	-0.06	1.33	98.1
Market Timing	0.93	0.04	0.03	0.99	100.0
Managed Futures	1.13	0.01	-0.15	1.47	99.1
Merger Arb.	0.83	0.11	0.06	0.77	100.0
Relative Value	0.89	0.10	0.01	0.96	91.9
Sector	1.01	0.06	-0.07	1.26	99.0
Short Selling	1.00	0.07	-0.07	1.05	100.0
All					98.2

This table presents the estimates presents the average values of θ_0, θ_1, θ_2 and ξ for each strategy in the constrained MA(2) model. The last column presents the percentage of funds for which we can obtain possible solutions

We can see that constraining the GLM model to be an MA(2) could lead to unsatisfactory results. Indeed, for some strategies, we have negative values (weights) for θ_1 and θ_2 overall and the consequences are less desirable for the most liquid strategies. This is especially true for Equity market neutral, Macro, Managed futures, Short selling and Fixed income diversified, which have, on average, a value of θ_0 greater than or equal to one. GLM (2004) obtained similar results for some strategies in their database[7]. This leads to a smoothing index of $\xi > 1$ and in turn, a lower volatility of unsmoothed returns, which is contrary to the model's hypothesis. For those strategies, the unsmoothing process will then lead to an underestimation of the funds' risk-adjusted performance. However, for more illiquid strategies, imposing an MA(2) model does not necessarily raise this problem. The average value of θ_0 for Convertible arbitrage, Distress securities, Fixed income convertible bonds, Fixed income high yield, and Fixed income mortgage is less than one and their smoothing index is also less than one.

In table 5, we present the results for the second approach in which we do not constrain the model to be an MA(2). Column six shows for each strategy, the percentage of funds exhibiting no statistically significant serial correlation. Column seven shows the percentage for which only the first-order serial correlation is statistically significant and column eight shows the percentage for which both the first- and the second-order serial correlation are statistically significant.

The unsmoothing process is then applied to each fund on a case-per-case basis. We recall that one of the objectives of this study is to compare the pure persistence of hedge funds across all strategies for smoothed and unsmoothed returns. Therefore, we should have the same number of funds when comparing the smoothed and unsmoothed returns of a strategy, and when it is not possible to unsmooth a fund's returns, the fund is excluded. Fortunately, as can be seen, we did not exclude many funds; of the 4,783 funds in the sample, we only excluded 1.8%. The percentage of exclusion differs of course from strategy to strategy; it is more than 10% for Fixed income diversified only (10.8%), but the strategy's weight in the sample is not of great significance. Only 7 funds were excluded from this strategy.

As can be seen in column six, it is not necessary to unsmooth returns for the majority of funds for liquid strategies. Indeed, for Equity hedge, Equity market neutral, Equity non hedge, Macro, Managed futures, Sector and Short selling, at least 80% of funds do not need to be unsmoothed as their serial correlation is not statistically significant. This is not the case for illiquid strategies where Convertible arbitrage, Fixed income convertible bonds and Fixed income mortgage exhibit a significant percentage of funds which must be unsmoothed up to lag 2. It can also be seen that with the unconstrained model, we always obtain satisfactory solutions as it takes into account the fund's level of serial correlation. It is also interesting to notice that for Macro, Managed futures and Short selling funds there is no need to unsmooth returns up to lag 2.

5.2.1 Robustness check for discontinuity

Figure A.1 and A.2 (in the appendix) show the histograms for reported and unsmoothed returns, and the value of the test statistic of the bins bracketing zero. This t-test measures whether the

[7] They used returns of 909 hedge funds from TASS database. The period of estimation starts from November 1977 to January 2001. HFR and TASS database don't have the same classification for hedge funds, but in their study, Equity hedge, Macro, Managed futures and Short selling are among strategies that exhibit a value of θ_0 higher to one and/or negative values for θ_1 or θ_2.

Table 5: Serial Correlation: Unconstrained model

	θ_0	θ_1	θ_2	ξ	% of Funds selected All	$k=0$	$k=1$	$k=2$
Convertible Arb	0.62	0.31	0.20	0.51	96.7	9.8	55.4	31.5
Distress Sec.	0.71	0.26	0.19	0.59	98.1	34.6	54.2	9.3
Emerging Mkt	0.75	0.22	0.19	0.63	99.5	75.5	20.4	3.6
Equity Hedge	0.72	0.23	0.21	0.59	98.8	79.7	14.4	4.6
Equity Mkt Neutral	0.75	0.23	0.20	0.62	95.9	79.3	14.5	2.1
Equity Non Hedge	0.75	0.24	0.22	0.63	98.3	80.2	17.4	0.8
Event Driven	0.73	0.25	0.16	0.61	99.4	51.7	40.8	6.9
FI Arbitrage	0.65	0.25	0.24	0.54	94.3	64.3	18.6	11.4
FI Convertible	0.73	0.20	0.19	0.60	100.0	42.9	38.1	19.0
FI Diversified	0.68	0.28	0.15	0.56	89.2	69.2	15.4	4.6
FI High Yield	0.68	0.30	0.17	0.57	100.0	44.0	52.0	4.0
FI Mortgage	0.60	0.25	0.25	0.49	94.7	55.3	15.8	23.7
Fund of Funds	0.73	0.24	0.18	0.61	99.3	70.2	25.2	3.9
Macro	0.77	0.23	NaN	0.65	96.7	86. 3	10.4	0.0
Market Timing	0.64	0.19	0.20	0.49	100.0	75.0	4.2	20.8
Managed Futures	0.79	0.21	NaN	0.68	96.9	91.9	4.9	0.0
Merger Arb.	0.70	0.23	0.28	0.58	100.0	62.8	27.9	9.3
Relative Value	0.68	0.28	0.20	0.57	93.8	54.5	31.8	7.6
Sector	0.70	0.23	0.22	0.58	98.4	83.8	9.9	4.7
Short Selling	0.78	0.22	NaN	0.66	100.0	84. 6	15.4	0.0
All								98.2

This table presents the average estimates of θ_0, θ_1, θ_2 and ξ for each strategy in the unconstrained approach. Column five presents the percentage of funds for which we can obtain possible solutions. Column six shows for each strategy, the percentage of funds exhibiting no statistically significant serial correlation. Column seven shows the percentage for which only the first-order serial correlation is statistically significant and column eight shows the percentage for which both the first- and the second-order serial correlation are statistically significant.

height of the vertical bar is different than that expected given the smoothed kernel estimate of the underlying distribution. The figures show a sharp discontinuity in the distribution of reported returns at zero, which can be interpreted as an underrepresentation of returns just below zero and overrepresentation just above zero. The discontinuity disappears in the distribution of unsmoothed returns which means that for these returns, the number of observations in each bin around zero is not statistically different than that expected. These results show that our unsmoothing procedure eliminates the concern regarding the presence of a discontinuity in the distribution of hedge fund returns. These findings should not be surprising because there is a positive relationship between discontinuity and illiquidity, the latter creating the serial correlation in returns; and the very purpose of the unsmoothing is to remove this serial correlation. Therefore, the unsmoothed returns should not exhibit any discontinuity at zero.

Nonetheless, our sample of true returns is not necessarily free from discontinuity because we didn't unsmooth all the returns namely those of funds that exhibit an insignificant serial correlation (k = 0). In this group, some funds exhibit a negative and insignificant serial correlation and their returns cannot be unsmoothed otherwise we will get incongruous results (negative θs), and others exhibit a positive and insignificant serial correlation. Figures B.1 and B.2 show the test of discontinuity for these two sub-groups. The figures show that for funds that exhibit a negative and insignificant serial correlation, the number of observations just below zero is not different than that expected but the number of observations just above zero seems to be over-represented. While for funds that exhibit a positive and insignificant serial correlation, there is a discontinuity just below and just above zero. However, the discontinuity for those funds remains less pronounced than that of reported returns of funds exhibiting a statistically significant serial correlation. Even if the returns of this second sub-group do not exhibit a significant serial correlation, they can also be unsmoothed in order to clear the discontinuity.

It is also important to mention that our persistence analysis before and after unsmoothing accounts for the possibility of distorted returns in order to avoid reporting losses because null returns are considered as negative returns in the estimation of transition probabilities[8].

5.3 Persistence of hedge fund returns

Table 6 compares for each strategy, the average positive persistence for funds with no serial correlation and those for which it is necessary to unsmooth returns. Columns 3 and 5 show the proportion of funds that exhibit a statistically significant positive persistence at the 5% level[9]. We can note that, on average, funds with smoothed returns have a higher level of positive persistence (except for Emerging market), and the difference may be significant. We also note that there are more funds exhibiting a statistically significant positive persistence in the universe of smoothed returns funds than in the universe of non-smoothed returns funds. These results suggest that the smoothing of returns may contribute to an increase in positive persistence. It is nevertheless important to notice that the majority of funds of nearly all strategies (to the exclusion of Managed futures and Short selling) exhibit statistically significant positive persistence for both smoothed and unsmoothed returns.

To verify whether smoothing contributes to an increase in the positive persistence of returns, we evaluated the persistence of smoothed and unsmoothed returns of funds exhibiting a statistically significant serial correlation of returns.

The results are presented in table 7 where we observe that for these funds, the average positive persistence drops considerably when one unsmooths the returns. The average drop of positive persistence across all strategies ranges from -9.1% for Market timing to -25.4% for Short selling, even if the persistence is not statistically significant for any fund of the latter. We also observe, across all strategies, a decrease in the percentage of funds exhibiting a statistically significant positive persistence at the 5% level. Distress securities, Fixed income high yield, Fixed

[8]The indicator variable I_t takes the value 0 if the return $R_t \leq 0$ and the value 1 if $R_t > 0$.

[9]The persistence is statistically significant for a fund at the 5% level if the statistic $t = \frac{\hat{p}_{11} - 0.5}{1/\sqrt{2n}} > 1.645$, where n is the number of monthly returns for that fund.

Table 6: Positive persistence for funds with no serial correlation and for funds with first- or second-order serial correlation

	Funds with k=0		Funds with k=1 or k=2	
Strategy	p_{11}	% Sign.	p_{11}	% Sign.
Convertible Arb	0.73	88.9	0.84	98.8
Distress Sec.	0.79	89.2	0.84	100.0
Emerging Mkt	0.73	83.1	0.73	97.9
Equity Hedge	0.67	70.3	0.72	91.5
Equity Mkt Neutral	0.67	62.1	0.72	87.5
Equity Non Hedge	0.66	67.0	0.72	95.5
Event Driven	0.73	85.6	0.80	96.4
FI Arbitrage	0.77	77.8	0.80	85.7
FI Convertible	0.64	55.6	0.66	50.0
FI Diversified	0.73	80.0	0.85	100.0
FI High Yield	0.84	100.0	0.85	100.0
FI Mortgage	0.84	95.2	0.91	100.0
Fund of Funds	0.75	90.5	0.80	99.4
Macro	0.65	54.1	0.68	86.4
Market Timing	0.62	66.7	0.87	100.0
Managed Futures	0.59	26.8	0.60	36.4
Merger Arb.	0.77	92.6	0.82	100.0
Relative Value	0.76	82.6	0.84	98.8
Sector	0.68	61.9	0.72	89.3
Short Selling	0.55	0.0	0.62	0.0

This table presents the average positive persistence for funds with no serial correlation and those for which it is necessary to unsmooth returns. Columns 3 and 5 show the proportion of funds that exhibit a statistically significant positive persistence at the 5% level.

income mortgage and Funds of funds exhibit the highest proportion of funds with a statistically significant positive persistence. Managed futures, Macro and Short selling have the lowest proportion of funds with statistically significant positive persistence. Another important point to mention here is that for almost all strategies, the average positive persistence of unsmoothed returns for funds with $k = 1$ or 2, ends up being lower than the average positive persistence for funds with no serial correlation (Table 6). The exception comes from Fixed income diversified (0.75 vs. 0.73) and Market timing (0.79 vs. 0.62).

Overall, our findings suggest that the smoothing of returns (voluntary or involuntary) is done at the advantage of the manager given that it contributes to an increase in the persistence of his positive returns.

If we aggregate the positive persistence of returns for funds with no serial correlation and the positive persistence of unsmoothed returns for funds with serial correlation, we obtain

Table 7: Positive persistence of smoothed and unsmoothed returns for funds with k=1 or k=2

	Smoothed returns		Unsmoothed returns		
	p_{11}	% Sign.	p_{11}	% Sign.	% Var. of p_{11}
Convertible Arb	0.84	98.8	0.66	67.50	-20.8
Distress Sec.	0.84	100.0	0.71	80.88	-15.0
Emerging Mkt	0.73	97.9	0.63	61.70	-12.8
Equity Hedge	0.72	91.5	0.60	42.86	-16.8
Equity Mkt Neutral	0.72	87.5	0.60	50.00	-17.1
Equity Non Hedge	0.72	95.5	0.62	45.45	-13.6
Event Driven	0.80	96.4	0.68	78.31	-14.5
FI Arbitrage	0.80	85.7	0.67	66.67	-16.9
FI Convertible	0.66	50.0	0.55	25.00	-15.9
FI Diversified	0.85	100.0	0.75	76.92	-12.4
FI High Yield	0.85	100.0	0.69	82.14	-18.7
FI Mortgage	0.91	100.0	0.78	93.33	-14.7
Fund of Funds	0.80	99.4	0.71	85.04	-11.5
Macro	0.68	86.4	0.55	36.36	-18.8
Market Timing	0.87	100.0	0.79	100.00	-9.1
Managed Futures	0.60	36.4	0.51	9.09	-14.6
Merger Arb.	0.82	100.0	0.69	81.25	-15.9
Relative Value	0.84	98.8	0.70	68.67	-16.3
Sector	0.72	89.3	0.62	53.57	-13.8
Short Selling	0.62	0.0	0.46	0.00	-25.4

This table presents the average positive persistence for smoothed and unsmoothed returns of funds exhibiting a statistically significant serial correlation of returns. Columns 3 and 5 show the proportion of funds that exhibit a statistically significant positive persistence at the 5% level. Column 6 measure the variation between the persistence measure for smoothed and unsmoothed returns.

the results presented in table 8, which represent the average "true" positive persistence for each strategy. With aggregate data, the majority of funds for most strategies exhibit statistically significant positive persistence at the 5% level (17 out of 20 strategies). At the 1% level, it is the case for 9 strategies of which arbitrage strategies, fixed income strategies, FOF and other strategies based on illiquid securities (Convertible arbitrage, Distress securities, Event driven, Fixed income arbitrage, Fixed income high yield, Fixed income mortgage, FOF, Merger arbitrage and Relative value arbitrage). The lowest values of persistence are for Short selling (0.54), Managed futures (0.58) and Fixed income convertible bonds (0.59) and the highest are for Fixed income mortgage (0.82), Fixed income high yield (0.76) and some arbitrage strategies.

Table 8: Positive persistence of true returns for all funds

	p_{11}	% Sign. at 5%	% Sign. at 1%
Convertible Arb	0.67	69.7	51.7
Distress Sec.	0.74	83.8	72.4
Emerging Mkt	0.70	77.9	44.1
Equity Hedge	0.66	65.0	35.7
Equity Mkt Neutral	0.66	60.0	34.1
Equity Non Hedge	0.65	63.0	41.2
Event Driven	0.71	82.1	63.0
FI Arbitrage	0.74	74.2	66.7
FI Convertible	0.59	38.1	23.8
FI Diversified	0.73	79.3	50.0
FI High Yield	0.76	90.0	72.0
FI Mortgage	0.82	94.4	91.7
Fund of Funds	0.74	88.9	71.0
Macro	0.64	52.2	28.8
Market Timing	0.67	75.0	45.8
Managed Futures	0.58	25.9	11.6
Merger Arb.	0.74	88.4	79.1
Relative Value	0.74	76.8	64.6
Sector	0.67	60.6	35.1
Short Selling	0.54	0.0	0.0

This table presents the positive persistence of returns for funds with no serial correlation and the positive persistence of unsmoothed returns for funds with serial correlation. Columns 3 and 4 present the proportion of funds that exhibit a statistically significant positive persistence at the 5% and 1% levels respectively.

5.4 Persistence vs. probability of positive returns

Positive persistence evaluates a manager's ability to deliver consecutive positive returns. The approach focuses on each past positive return and observes the sign of the following one. Although this information is relevant, it does not necessarily provide insight as to the odds of delivering positive or negative returns. For that purpose, we should estimate the unconditional probability of positive returns, P_1, which takes into account the number of positive returns during the evaluation period. To support our assertion, let us consider the following example. Suppose a manager whose performance over 10 periods is as follows, where 1 represents the occurrence of a positive return and 0 that of a non-positive return:

$$0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 0$$

The probability of positive returns and the positive persistence can be estimated by counting, respectively, the number of 1s and the number of subsequent 1s. In this case, $P_1 = 3/10 = 0.3$, and $p_{11} = 2/3 = 0.66$. This can be interpreted as a positive persistence, but a poor performance on a regular basis[10] (low value of P_1). However, looking only at p_{11} is misleading when evaluating the manager's overall performance. Another look at this example shows that there is also the presence of negative persistence. In fact, if there is positive persistence and negative persistence, a high value of p_{11} will not be an indication of a high probability of positive returns. On the other hand, if there is positive persistence and no negative persistence, the values of p_{11} and P_1 should not be very different and a high positive persistence will be an indication of a high probability of positive returns. Table 9 presents the average values of p_{11}, p_{00} and P_1 for the hedge fund strategies.

Table 9: Positive/negative persistence and the probability of positive/negative returns

	p_{11}	% Sign.	p_{00}	% Sign.	P_1	% Sign.
Convertible Arb	0.67	69.7	0.32	0.00	0.68	78.65
Distress Sec.	0.74	83.8	0.30	0.00	0.73	95.24
Emerging Mkt	0.70	77.9	0.33	0.00	0.69	82.56
Equity Hedge	0.66	65.0	0.37	0.20	0.65	70.00
Equity Mkt Neutral	0.66	60.0	0.36	0.00	0.65	67.03
Equity Non Hedge	0.65	63.0	0.41	1.68	0.63	62.18
Event Driven	0.71	82.1	0.33	0.58	0.70	84.97
FI Arbitrage	0.74	74.2	0.28	0.00	0.74	86.36
Convertible	0.59	38.1	0.41	0.00	0.59	47.62
FI Diversified	0.73	79.3	0.33	0.00	0.72	86.21
FI High Yield	0.76	90.0	0.30	0.00	0.75	96.00
FI Mortgage	0.82	94.4	0.26	2.78	0.81	97.22
Fund of Funds	0.74	88.9	0.36	0.23	0.71	88.81
Macro	0.64	52.2	0.37	0.00	0.64	65.37
Market Timing	0.67	75.0	0.37	0.00	0.65	70.83
Managed Futures	0.58	25.9	0.43	1.39	0.58	27.31
Merger Arb.	0.74	88.4	0.25	0.00	0.75	100.00
Relative Value	0.74	76.8	0.29	0.51	0.73	83.84
Sector	0.67	60.6	0.37	0.00	0.66	70.21
Short Selling	0.54	0.0	0.56	15.38	0.49	0.00

This table contrasts positive persistence and the probability of positive returns. p_{11} is the measure of positive persistence, p_{00} is the measure of negative persistence and P_1 is the unconditional probability of positive returns. Columns 3, 5 and 7 present the proportion of funds for which the parameter estimates are statistically significant at the 5%.

We can see that for almost all strategies there is no negative persistence except for Short selling funds of which about 15% of funds (2 out of 13) have a statistically significant value of $p_{00} > 0.5$ at the 5% level. This means that a monthly loss is generally followed by a gain in the hedge fund's universe. Column 6 shows the probability of positive returns. We can see that in general, the values of p_{11} are not very different from those of P_1; this is due to the absence of

[10]Here, we don't take into account the level of returns.

negative persistence of returns in the hedge fund's universe. The last column shows the percentage of funds for which the probability of positive returns is statistically superior to 0.5 at the 5% level[11]. We can note that except for Short selling, Managed futures and Fixed income convertible bonds, the majority of funds have a probability of positive returns superior to 0.5. The highest proportion is recorded for Merger arbitrage where all the funds present a statistically significant probability of delivering positive returns. It is followed by Fixed income mortgage (97.22%) and Fixed income high yield (96%).

On the basis of these results, we can conclude that despite a context where markets have been faced with difficult periods since the year 2000, hedge funds have been able to deliver positive returns and have done so in a sustainable manner until 2007. Arbitrage strategies, and some fixed income strategies, seem to be more prone to deliver absolute returns.

5.5 Persistence with respect to a high watermark

Although the results obtained above are interesting, the measures used unfortunately don't account for the level of returns. It is important to note that the absence of negative persistence in hedge fund returns (p_{00} not statistically superior to 0.5), doesn't mean that in the case of a loss, the capital will be recovered during the next period (month), but simply that after a loss there is a strong probability that the return will be positive during the next month. It is important to seize the fact that the fund's capacity to recover losses in the subsequent period depends on both the size of the loss and the manager's ability to generate positive returns of the same amplitude. Therefore, when a fund experiences a large drawdown, it will require a significant profit in the subsequent period, or a series of small profits, to recover the capital lost. This aspect is relevant for hedge funds because it is well known that several strategies, in particular arbitrage strategies, tend to generate positive returns of small amplitude, but when they face losses, the latter are often larger in amplitude. It can often take several periods for a fund to recover lost capital. Taking into account the level of returns also gives an indication as to the evolution of a manager's high water mark. Most hedge funds are subject to a high water mark criterion, which means that the manager will only receive performance fees on that particular pool of invested money when its value exceeds its previous maximum value. By accounting for the level of returns, we can estimate the "performance with respect to the high water mark", which can be defined as the probability of increasing the high water mark, and the "persistence with respect to the high

[11] To get those values we calculate the confidence interval of $\hat{P}_{1,i}$ for each fund i, with

$$\hat{P}_{1,i} = n_{+,i}/n_i$$

where $n_{+,i}$ = number of positive returns and n_i is the size of sample for fund i. By the central limit theorem, we have:

$$\Pr\left[(n_{+,i}/n_i) - 1.96\sqrt{\frac{(n_{+,i}/n_i)(1-(n_{+,i}/n_i))}{n_i}} < P_{1,i} < (n_{+,i}/n_i) + 1.96\sqrt{\frac{(n_{+,i}/n_i)(1-(n_{+,i}/n_i))}{n_i}}\right] = 0.95$$

If the lower bound of this confidence interval is superior to 0.5, the probability is statistically superior to 0.5 at 5% level. We can notice that the smaller n_i is, the larger the confidence interval is.

water mark", which in turn can be defined as the probability of increasing the high water mark during the next period given that it has been increased during the current period. These estimations are performance measures in the sense that they give the frequency at which a manager is able to receive performance fees[12].

Let us define C_t and H_t respectively as the cumulative wealth and the high water mark at time t. Then:

$$\begin{aligned} C_t &= C_{t-1}(1+r_t) \\ H_t &= \max(C_t, H_{t-1}) \end{aligned} \tag{41}$$

with C_0 and H_0 normalized at \$1.

Let us also define the dichotomous variable I'_t which takes the following values:

$$\begin{aligned} I'_t &= 1 \;\; if \;\; H_t > H_{t-1} \\ I'_t &= 0 \;\; if \;\; H_t = H_{t-1} \end{aligned} \tag{42}$$

By this process, and as in the preceding model, we can obtain the following probabilities:

$$\begin{aligned} P'_1 &= \Pr[I'_t = 1] \\ p'_{11} &= \Pr[I'_{t+1} = 1 | I'_t = 1] \\ p'_{00} &= \Pr[I'_{t+1} = 0 | I'_t = 0] \end{aligned}$$

P'_1 is the probability of increasing the high water mark and it could also be defined as the probability of receiving performance fees. p'_{11} is the persistence in increasing the high water mark or the probability of increasing the high water mark during the next period given that it has been increased during the current period; it could also be defined as the persistence of the receipt of performance fees. p'_{00} is the persistence of the stagnation of the high water mark, or the probability of having the same high water mark for the next period since it did not change for the current period; it could also be defined as the persistence of the absence of performance fees.

It is important to note that these measures are settled for an investment made at the fund's inception date and enable us to compare all of the funds on the basis of their performance since inception. Indeed, a manager will have a different high water mark for each investment made at a different time. Therefore, when $I'_t = 1$, it means that the manager receives performance fees from an investor who invested money at time $t = 0$ and when $I'_t = 0$, it means that he does not receive any performance from an investor who invested at that time. However, $I'_t = 1$ means that the manager receives performance fees from all investors who have invested from time *0* to time *t-1*, and $I'_t = 0$ means that he does not does not receive performance fees from all investors who have invested during this period, but he could receive performance fees from

[12] Here, we assume a hurdle rate of 0% given that it is the value generally applied by hedge fund managers.

certain investors who have invested between time *1* and time *t-1*.

Table 10 shows the values of the three probabilities for each strategy. First of all, we can see that the values of p'_{11} are not so different from those of p_{11} (table 9) (this could be interpreted as a similarity in the persistence of positive returns and the persistence of the receipt of performance fees). But this doesn't mean that measuring persistence based on returns is equivalent to measuring persistence with respect to a high water mark.

Table 10: Persistence with respect to the high water mark

	p'_{11}	% Sign.	p'_{00}	% Sign.	P'_1	% Sign.
Convertible Arb	0.68	74.2	0.69	75.3	0.48	18.0
Distress Sec.	0.74	82.9	0.65	59.0	0.57	35.2
Emerging Mkt	0.69	70.8	0.72	67.2	0.46	16.4
Equity Hedge	0.66	63.8	0.75	79.4	0.41	9.2
Equity Mkt Neutral	0.67	62.7	0.73	73.0	0.43	11.4
Equity Non Hedge	0.64	60.5	0.79	87.4	0.36	1.7
Event Driven	0.71	82.1	0.69	72.8	0.50	23.1
FI Arbitrage	0.74	78.8	0.62	45.5	0.60	42.4
Convertible	0.56	42.9	0.83	95.2	0.28	4.8
FI Diversified	0.75	81.0	0.69	74.1	0.54	29.3
FI High Yield	0.76	90.0	0.64	58.0	0.61	48.0
FI Mortgage	0.83	94.4	0.62	47.2	0.66	75.0
Fund of Funds	0.73	84.4	0.71	73.8	0.51	23.8
Macro	0.64	53.2	0.78	88.8	0.38	3.9
Market Timing	0.62	58.3	0.75	87.5	0.40	16.7
Managed Futures	0.54	29.6	0.84	94.4	0.26	2.3
Merger Arb.	0.76	93.0	0.66	69.8	0.58	44.2
Relative Value	0.75	80.3	0.65	60.6	0.57	43.9
Sector	0.66	58.5	0.74	78.7	0.42	11.7
Short Selling	0.53	23.1	0.97	100.0	0.06	0.0

This table presents the positive and negative persistence with respect to the high water mark and probability of increasing the high water mark true returns

In fact, when we look at the values of p'_{00} and P'_1, we see that they are different from those of p_{00} and P_1. Whereas the probability that a loss will be followed by another loss is low for all strategies, the probability that a non-payment of performance fees from all investors will be followed by another non-payment of performance fees is generally high. The fact that the values of p'_{11} and p'_{00} are statistically superior to 0.5 for the majority of funds in almost all strategies means that when a manager receives performance fees for a given period, there is a high probability that he will receive performance fees for the next period; but it also means that when he doesn't receive performance fees for a given period, there also a high probability that he will not receive performance fees for the next period because he will not be able to recover the capital lost during that period. These results are in line with what we stated previously, i.e. hedge fund managers generally generate positive returns of small amplitude, but when they face losses, the latter are often of larger amplitude and the managers are unable to rapidly recover the capital lost. In terms of persistence, this translates into small high water mark increases during good periods and a stagnation of the high water mark after a bad period.

Another important point is that strategies for which there is a higher persistence in the increase of the high water mark are those where there is a lower persistence in the stagnation of the high water mark, notably for Fixed income mortgage (0.83 vs. 0.62), Fixed income high yield (0.76 vs. 0.64) and Merger arbitrage (0.76 vs. 0.66). And vice versa, notably for Short selling (0.53 vs. 0.97), Managed futures (0.54 vs. 0.84) and Fixed income convertible bonds (0.56 vs. 0.83). These results seem to show that for strategies such as Merger arbitrage and others that exhibit a higher positive persistence with respect to a high water mark, managers show more ability to bring the capital back to a value superior or equal to that preceding the loss. Is it because they have superior skills? It is difficult to answer this question. However, we can address the question as to whether these strategies exhibit a shorter time for the recovery of lost capital. For this purpose, we must perform a thorough examination given that the previous estimations concern the high water marks from all investments. Therefore, certain investments made at different moments could be recovered at a certain time but not others, and the " overall" high water mark will not increase, thus resulting in the stagnation of the "overall" high water mark. For instance, for an investment made at the inception date, the manager may have a certain high water mark at time t and receive a new investment at the end of time $t + 1$. At that moment, he could face a loss followed by another loss at time $t + 2$ and a gain at time $t + 3$. The capital of the investment made at time $t + 1$ could be recovered at time $t + 3$ and the manager could receive performance fees from this investment, but this doesn't necessarily mean that he will also recover the capital lost at time $t + 1$ (from the investment made at the inception date). Therefore, a better way of gauging the ability to recover capital after a loss is to estimate the recovery time for each loss and observe the average for each fund and each strategy. This issue will be addressed in the next section.

Column 6 of table 10 presents the unconditional probability of increasing the high water mark. We can see that contrary to the results of table 9, p'_{11} and P'_1 are not similar as p_{11} and P_1 were; this is due to the high values of p'_{00} that indicate an inverse persistence. The relatively low values of P'_1 for the majority of strategies (they are inferior to 0.60 for 17 out of 20 strategies) mean that the managers are unable to increase their high water mark on a regular basis even though they have, in general, a high probability of delivering positive returns (see values of P_1). Nevertheless, Fixed income mortgage and Fixed income high yield managers are more prone to increase their high water mark, whereas Short selling, Managed futures and Fixed income convertible bonds managers are less prone to increase their high water mark.

Based on these results, we can state that even if hedge funds are able to deliver positive (absolute) returns, they have greater difficulty in increasing their high water mark on a regular basis. Indeed, periods of small, consecutive increases in the high water mark are often interrupted by periods of stagnation of the high water mark, which is due to their risk exposure that can lead to important drawdowns during bad periods. It is important to note that when we use the words “important drawdowns” it is does not mean that hedge funds hold high-risk positions leading to large losses, but simply that losses can be significant in comparison to gains. The specific risk-return profile of many hedge fund strategies characterized by payoffs similar to those of short puts on market indices has been mentioned by several studies (Fung and Hsieh (1997), Mitchell and Pulvino (2001), Agarwal and Naik (2004)). This option-like payoff can be modeled via a covered call[13]. A covered call is a strategy in which an investor writes a call option contract while at the same time owning an equivalent number of shares of the underlying stock. While this strategy can offer limited protection against a decline in the price of the underlying stock and limited profit participation with an increase in the stock price, it generates income because the investor keeps the premium received from writing the call. Thus, the investor will have a profit lower than that of the underlying stock if the latter increases substantially (the option will be exercised) and will have lower losses than the underlying stock. We are not implying that numerous hedge funds use covered call strategies, but using this kind of strategy can result in a payoff similar to that of hedge funds. Indeed, looking at the historical performance of certain hedge fund strategies, we observe that they have payoffs similar to that of a covered call on the S&P500 index, i.e. positive returns are generally small and losses are lower than those of the S&P500. A covered call can help a manager who aims to provide absolute returns because although it limits gains, it can contribute to increasing the regularity of these gains. Absolute returns dot not necessarily mean high returns, but ”good” returns, regardless of the market’s direction. On the other hand, even if the strategy helps to reduce losses the latter could be substantial in comparison with gains as the effect on gains and losses is not symmetrical and this results in a payoff with important drawdowns in comparison to gains.

5.6 Average time to recover capital after a loss

The previous results showed that when a hedge fund manager faces a loss, it could take a certain time before he recovers the capital. The measures of performance and persistence with respect to a high water mark provide not only an indication as to a manager’s performance, but also an indication as to the risk an investor could face when he invests in a hedge fund. Indeed, if an investor plans to withdraw his money after a loss, he should know that for the following period, there is a slight probability that the manager will bring the fund back to a level superior or equal to that preceding the loss. The investor should therefore wait a certain time if he wishes to withdraw an amount of capital superior or equal to the manager’s last high water mark. Then, in order to evaluate the right time to withdraw his money, he should take this aspect into account and also be aware of how much notice is required as this varies from one fund to another. We estimated the average time to recover capital after a loss on the basis of the previous results. For this purpose, for each loss recovered, we calculated the number of months necessary to return to a level of capital superior or equal to that preceding the loss. Table 11 shows the results for all strategies. The values are averaged for each fund and thereafter averaged for each strategy. Columns 2 to 5 show, respectively, the mean, the 25th percentile, the 75th percentile and the volatility of the average recovery time per strategy. Column 6 shows the average proportion of

[13]We take the example of a covered call in order to have a strategy that combines the trading of assets and derivatives given that hedge funds are general invested in traditional assets (stocks, bonds, etc.) as well as derivatives.

losses for each strategy. This statistic demonstrates the frequency of losses per strategy on the same basis given that the funds of each of the strategies do not have the same lifespan.

Table 11: Average time taken to recover capital after a drawdown (in months)

	Mean	25 prcnt	75 prcnt	Volatility	Mean loss	# of funds
Convertible Arb	3.99	2.87	4.75	1.61	32.24	89
Distress Sec.	3.00	2.17	3.37	1.44	27.00	105
Emerging Mkt	4.73	2.36	5.21	3.65	30.20	195
Equity Hedge	4.17	2.54	4.76	2.57	34.68	980
Equity Mkt Neutral	3.66	2.32	4.22	2.03	34.14	185
Equity Non Hedge	5.00	2.81	5.77	3.17	36.92	119
Event Driven	3.40	2.18	3.92	2.12	30.25	173
FI Arbitrage	2.60	1.73	3.31	1.27	25.26	66
Convertible	7.15	3.46	10.41	4.31	40.95	21
FI Diversified	3.03	1.97	3.97	1.45	27.65	58
FI High Yield	2.49	1.75	2.88	1.03	24.70	50
FI Mortgage	4.14	1.75	6.73	3.19	18.90	36
Fund of Funds	3.43	2.43	3.74	1.85	28.65	1734
Macro	4.25	2.91	4.87	2.23	36.04	205
Market Timing	5.21	2.81	7.09	3.37	34.72	24
Managed Futures	4.63	3.46	5.46	1.85	41.84	216
Merger Arb.	3.18	2.33	4.04	1.24	25.05	43
Relative Value	2.94	1.98	3.38	1.63	26.74	198
Sector	4.14	2.41	5.00	2.32	34.09	188
Short Selling	7.12	5.73	8.10	2.33	50.57	13

We note that for hedge fund strategies, the average time to recover a capital loss is more than 3 months even though a manager may be able to recover the amount earlier for some strategies (Fixed income high yield (2.49), Fixed income arbitrage (2.60) and Relative value arbitrage (2.94)). This could be explained by the fact that either the managers of those strategies do not face great losses in general, or that they assume a high level of risk after a loss in order to recover the capital quickly. For other strategies it takes more time to recover the capital after a loss, notably for Fixed income convertible bonds (7.15), Short selling (7.12) and Market timing (5.21). One would expect that funds exhibiting higher positive persistence should take less time to recover the capital, but this is not necessarily the case. The negative relation between the time to recover the capital after a loss and the positive persistence seems to be more obvious for funds exhibiting lower positive persistence of returns and with respect to the high water mark. Short selling, Fixed income convertible bonds and Managed futures funds are among those generally taking more time to recover capital after a loss. However, for funds exhibiting higher positive persistence the relation is only confirmed for Fixed income high yield. Merger arbitrage funds hold the sixth position in terms of time to recover capital and Fixed income mortgage funds, which exhibit the highest positive persistence of returns and with respect of a high water mark, hold the tenth position. One reason could be that Fixed income mortgage funds exhibit negative outliers, more so than other fund categories (they exhibit the lowest (-1.66) asymmetry and the highest kurtosis (20.46)). Concerning the proportion of losses, hedge funds generally exhibit fewer losses than gains, except for Short selling for which the number of losses and gains is almost the same. The lowest proportions are generally attributed to Fixed income strategies and Managed futures.

These results show how the advance notice imposed by most hedge funds constitutes not only effective protection against withdrawals from investors in need of liquidity, but also from unhappy investors following a loss. Indeed, given the asymmetry in the amplitude of gains and losses, it is important for a manager to set up some delay for withdrawals of money, especially investors attempting to withdraw after a loss. Advance notice that exceeds the average time to recover capital may enable the manager to bring the capital back to its pre-loss value, thus giving the investor time to change his mind. The fact that the average time to recover a loss is more than 3 months for most strategies suggests that a median advance notice of 30 days is not necessarily optimal for hedge fund managers. Managers with an advance notice in excess of 3 months will probably have a greater chance of retaining unhappy investors ready to withdraw their money after a loss. However, due to competition between managers, it may be difficult to establish long periods of advance notice, even if in the case of our data the maximum advance notice is one year.

Table 11 shows the average recovery time for losses that have been recovered. It is important to mention that in our sample some losses have not yet been recovered and have therefore been discarded from table 11. The unrecovered losses are not exclusively large losses, but also losses that occurred toward the end of our sample period. Table 12 exhibits the statistics for the unrecovered losses of each strategy. Column 2 presents the average proportion of unrecovered losses and Column 3 shows the average proportion of large losses among the unrecovered losses. Large losses are those for which the absolute value is higher than 2 standard deviations of the distribution of returns. We can see that Short selling and almost all fixed income strategies (Fixed income high yield, Fixed income diversified and Fixed income mortgage) are among those that exhibit the highest proportions of unrecovered losses. However, contrary to these fixed income strategies, the existence of unrecovered Short selling losses is not due, for a considerable proportion, to severe drawdowns. Indeed, the average proportion of large losses among unrecovered losses is only 1.92%, whereas this figures ranges from 15% to 24% for the relevant fixed income strategies. We can also see that losses are mostly recovered for Emerging market, Market timing and Merger arbitrage strategies.

Table 11 provides a good insight into the average time needed to recover a loss, but these results must be interpreted with caution given that for some strategies, such as fixed income strategies, many losses have not yet been recovered. Another interesting point is that the recovery period for losses can be quite significant. For example, the maximum recovery time exceeds 100 months for some managers (Equity hedge (115), Macro (114), Emerging market (113), FOF (111) and Equity non hedge (110)). We note that many of these drawdowns occurred between August '97 and July '98. In some cases, the individual losses were not very large. However, the managers were unable to generate sufficient subsequent positive returns prior to enduring another drawdown. The second half of 1998 was not a good period for the hedge fund industry,

Table 12: Statistics for unrecovered losses (in months)

	Proportion of of unrecovered losses	Proportion of unrecovered losses that were $> 2\sigma$
Convertible Arb	9.63	13.33
Distress Sec.	11.88	7.33
Emerging Mkt	5.60	5.78
Equity Hedge	9.79	10.35
Equity Mkt Neutral	10.96	12.36
Equity Non Hedge	9.92	5.37
Event Driven	10.76	13.42
FI Arbitrage	11.43	21.00
FI Convertible	9.95	4.37
FI Diversified	13.50	18.62
FI High Yield	16.76	23.67
FI Mortgage	11.93	15.37
Fund of Funds	8.51	11.83
Macro	8.77	7.63
Market Timing	6.05	8.65
Managed Futures	8.80	5.34
Merger Arb.	6.15	21.51
Relative Value	11.52	17.89
Sector	8.12	7.54
Short Selling	23.95	1.92

This table presents the statistics for unrecovered losses. Column 2 presents the average proportion of unrecovered losses and Column 3 shows the average proportion of large losses among the unrecovered losses. Large losses are those for which the absolute value is higher than 2 standard deviations of the distribution of returns.

or for the market in general, and it was thereby a difficult period for the recovery of prior losses. For instance, an investor who invested $1 in the particular Equity hedge fund that exhibited the longest recovery time would have waited for 115 months before breaking even. This also means that the manager would not have received any performance fees from this investor during those 115 months. However, this doesn't mean that the manager did not receive performance fees from other investors who entered the fund at a later date.

These findings do not augur well for investors in the forthcoming months. Indeed, given that the current financial crisis may have more negative impacts on the hedge fund industry than the 1998 crisis, one should expect that it may take a considerable time before investors recoup their losses.

6 Conclusion

In this study, we have addressed the issue of hedge fund performance persistence using a Markov chain model. Persistence is evaluated via transition probabilities, which make no a-priori as to the distribution of returns. Persistence is also evaluated after accounting for serial correlation in hedge fund returns, which is often due to the holding of illiquid assets or the manager's motivation to enhance his performance. For this purpose, we use a new approach based on the method of moments and on the model of Getmansky and al. (2004) to unsmooth returns. To assess the significance of persistence estimates, we also developed a t-test which accounts for the size of the sample of fund returns. Our approach also overcomes the issue of a "strategic" discontinuity in the return distribution around zero that Bollen and Pool (2009) identify and attribute to the fact that managers will adjust reported returns to minimize the chance of small negative returns in order to promote the appearance of 'pure persistence'.

Our study firstly shows that the unsmoothing of returns is not necessary for all funds, especially for those comprised of liquid strategies, namely Macro, Managed futures, Sector and Short selling funds. Therefore, imposing an MA(2) model for all funds as is the case for Getmansky and al. (2004) could lead to incongruous results. We also note that smoothing may contribute to an increase in the pure persistence of returns. Getmansky and al. (2004) have pointed out that the evidence of relative persistence found in some studies may be indirectly linked to serial correlation in returns. Our results show that for almost all strategies, the average positive persistence of returns of funds with no statistically significant serial correlation is lower than that of funds with smoothed returns; and the average persistence of the latter drops considerably (between -9.1% and -25.4%) when one unsmooths returns. Our findings nevertheless suggest that, to the exclusion of Short selling, Managed futures and Fixed income convertible bonds, the majority of funds of other strategies exhibit persistence of positive returns and almost all of the funds fail to exhibit persistence of negative returns. Our results show that until 2007, hedge funds were able to deliver sustained absolute returns despite periods of turbulence faced by the markets. We have yet to observe how the events of 2008 will affect these conclusions.

Hedge funds however exhibit difficulties in increasing their high water marks on a regular basis. Periods of consecutive positive returns are sometimes interrupted by large drawdowns which take several periods to recover because the positive returns are generally smaller in size. This translates into positive and negative persistence with respect to a high water mark. In other words, this leads to small and consecutive increases of the high water mark but also in stagnations of the high water mark over certain periods. The estimated average time to recover capital after a loss ranges from 2.49 months (Fixed income high yield) to 7.15 months (Fixed income convertible bonds). Given that the current financial crisis will no doubt intensify the negative asymmetry of the distribution of hedge fund returns, the average time to recover losses will increase, and with a median advance notice of 30 days, most of the funds will not have enough flexibility to reverse the situation in order to retain investors who are ready to withdraw their money. This will accentuate the liquidation of funds as it has been the case recently. Many analysts foresee that about one-third of hedge funds could be liquidated due to massive withdrawals on behalf of investors.

These results raise the question as to how an investor should evaluate a manager's performance, especially in terms of pure persistence. It is well known that the mean-variance analysis and the Sharpe ratio are not appropriate to evaluate the risk-adjusted performance of hedge funds because of the non-normal distribution of their returns. For the same reasons and in terms of pure persistence analysis, the persistence analysis with respect to the high watermark turns out to be a good alternative to the absolute persistence analysis (positive/negative returns). The persistence with respect to the high water mark provides a better way to account for the asymmetry between gains and losses and indicates the manager's ability to sustainably increase the investor's wealth because as long as the manager's high water mark fails to increase, the investor is no wealthier, even if the manager does deliver some positive returns.

References

[1] Agarwal, V., and N. Y. Naik, "Multi-Period Performance Persistence Analysis of Hedge Funds", *Journal of Financial and Quantitative Analysis*, September 2000 (2000b), vol. 35, pp. 327-342.

[2] Agarwal, V. and N. Naik, "Risks and Portfolio Decisions Involving Hedge Funds", *Review of Financial Studies*, Spring 2004, pp. 63-98.

[3] Asness, C., Krail, R. and J. Liew, "Do Hedge Funds Hedge?", *Journal of Portfolio Management,* 2001, 28(1), pp. 6-19.

[4] Bollen, C. and V. Pool,"Do Hedge Fund Managers Misreport Returns? Evidence from the Pooled Distribution ", *Journal of Finance*,Vol 64, No. 5, Octobre 2009, pp.2257-2258.

[5] Brooks, C. and H. M Kat, "The Statistical Properties of Hedge Fund Index Returns and their Implications for Investors", *Journal of Alternative Investments,* 2002, 5(2), pp. 26-44.

[6] Brown, S. and W. Goetzmann, "Performance Persistence", *Journal of Finance*, Vol. L, No 2, June 1995, pp. 679-698.

[7] Burgstahler, D. and I. Dichev. "Earnings Management to Avoid Earnings Decreases and Losses", *Journal of Accounting and Economics*, Vol. 24, 1997, pp.99-126.

[8] De Souza, C. and S. Gokcan, "Hedge Fund Investing: A Quantitative Approach to Hedge Fund Manager Selection and De-Selection", *Journal of Wealth Management*, Spring 2004.

[9] Fung, W. and D. Hsieh, "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds", *Review of Financial Studies*, Summer 1997, pp. 275-302.

[10] Geltner, D. "Estimating Market Values from Appraised Values without Assuming an Efficient Market", J*ournal of Real Estate Research,* 1993, 8, pp. 325-345.

[11] Getmansky, M., Lo, A.W. and I. Makarov, "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns", *Journal of Financial Economics,*2004, 74, pp. 529–609.

[12] Hasanhodzic, J. and A. Lo, “Can Hedge Fund Returns Be Replicated?: The Linear Case”, *Journal of Investment Management,* Vol. 5, No 2, 2007, pp. 5-45

[13] Jagannathan, R., A. Malakhov, and D. Novikov, “Do Hot Hands Exist Among Hedge Fund Managers? An Empirical Evaluation”, *NBER Working Papers 12015*, 2006.

[14] Kosowski, R., N. Naik, and M. Teo, “Do hedge funds deliver alpha? A Bayesian and bootstrap analysis”, *The Journal of Financial Economics*, 2007, Vol: 84, pp. 229-264

[15] Liang, Bing. “Hedge Funds: The Living and the Dead”, *Journal of Financial and Quantitative Analysis*, 2000, 35, pp. 309-326.

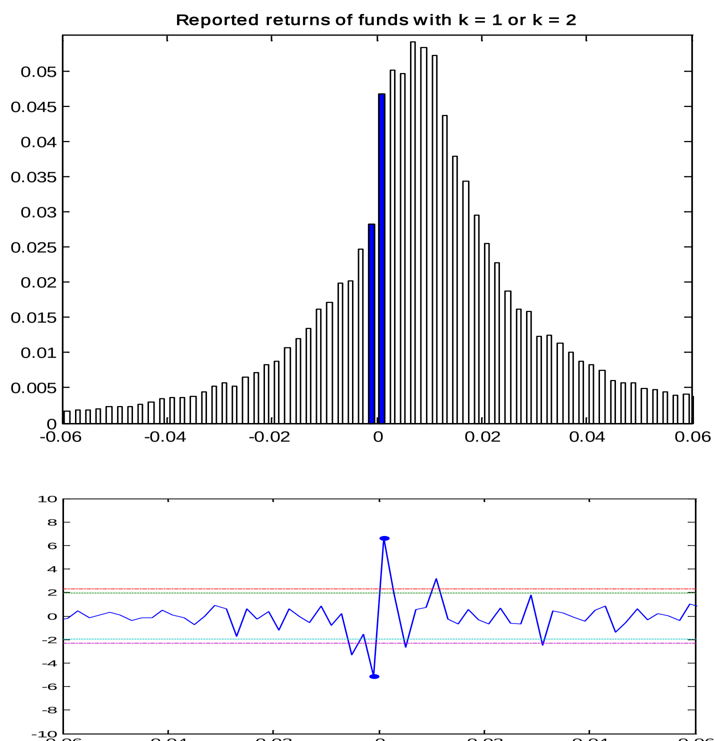
[16] Mitchell M. and T. Pulvino, “Characteristics of risk and return in risk arbitrage”, *Journal of Finance*, Vol. 56, 2001, pp. 2135-2175.

[17] Okunev, J. and D. White, “Hedge Fund Risk Factors and Value at Risk of Credit Trading Strategies”, *Working paper*, University of New South Wales, 2003.

Figure A: Test of discontinuity for funds with first or second order serial correlation (January 1994-December 2007)

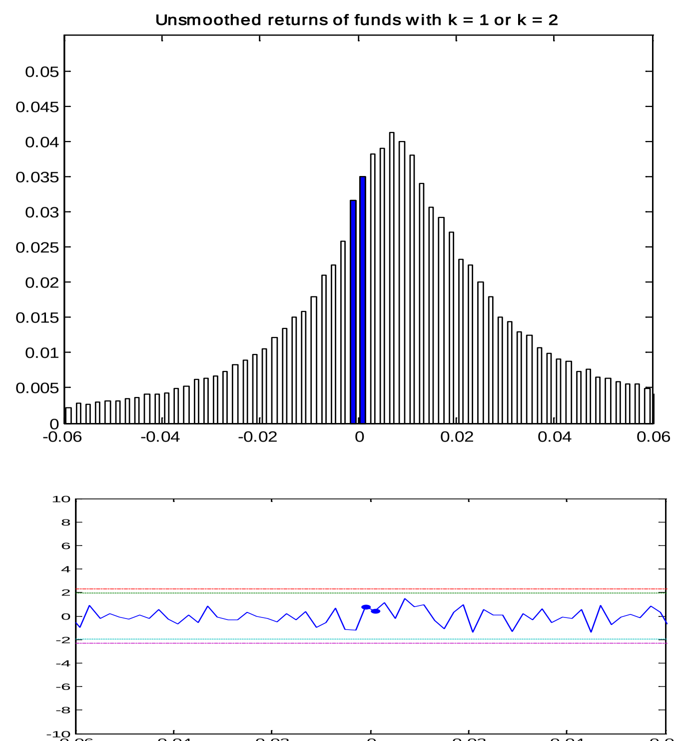
The top graphs display histograms of monthly reported and unsmoothed returns of funds that exhibit a statistically significant first and second order serial correlation. The size of each bin is 0.002 and bold bars indicate bins that bracket zero. The middle graphs show the value of the test statistics and the horizontal lines indicate 95% and 99% critical values. The bottom tables exhibit the test statistics of the six bins surrounding zero. The test statistics are distributed independent standard normal under the null hypothesis on no discontinuities in the histogram. We use the Gaussian Kernel to construct the reference distribution. For more information, see Bollen and Pool (2009). Fund of funds and Managed future funds are excluded for purpose of comparison their paper. Each fund in the sample is required to have at least 36 consecutive monthly returns.

Figure A.1



Bin	-2	-1	-0	+0	+1	+2
t-test	-3.28	-1.55	-5.13	6.62	1.88	-2.65

Figure A.2

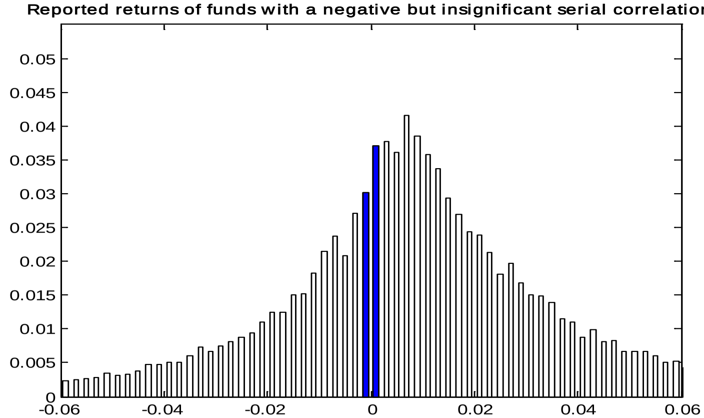


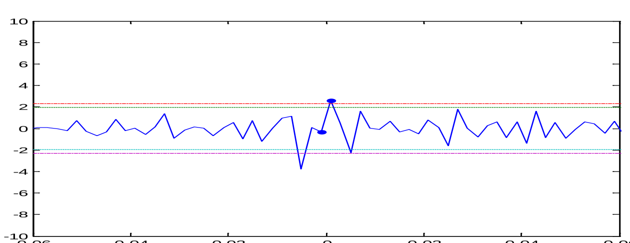
Bin	-2	-1	-0	+0	+1	+2
t-test	-1.11	-1.22	0.79	0.46	1.13	-0.18

Figure B: Test of discontinuity for funds with an insignificant serial correlation (January 1994-December 2007)

The top graphs display histograms of reported returns of funds that exhibit a negative or a positive insignificant serial correlation. The size of each bin is 0.002 and bold bars indicate bins that bracket zero. The middle graphs show the value of the test statistics and the horizontal lines indicate 95% and 99% critical values. The bottom tables exhibit the test statistics of the six bins surrounding zero. The test statistics are distributed independent standard normal under the null hypothesis on no discontinuities in the histogram. We use the Gaussian Kernel to construct the reference distribution. Fund of funds and Managed future funds are excluded for purpose of comparison their paper. Each fund in the sample is required to have at least 36 consecutive monthly returns.

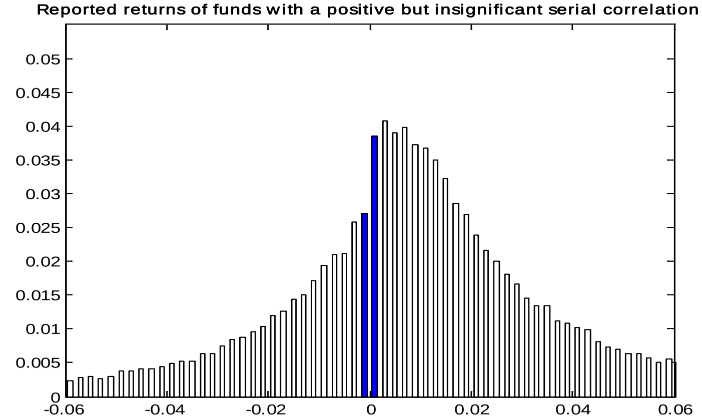
Figure B.1

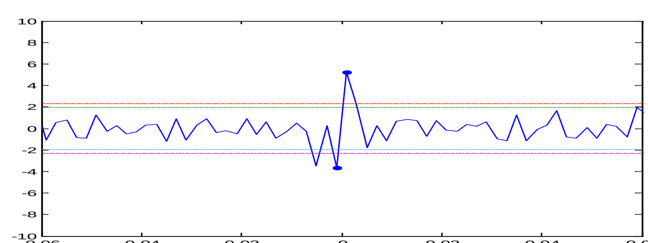




Bin	-2	-1	-0	+0	+1	+2
t-test	-3.74	0.06	-0.30	2.62	0.49	-2.24

Figure B.2





Bin	-2	-1	-0	+0	+1	+2
t-test	-3.46	0.26	-3.68	5.23	2.34	-1.77

A Estimation of $\sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)$ by Delta method

We know that $\hat{p}_{11}$ can also be expressed in the following way:

$$\hat{p}_{11}=\frac{\hat{P}_{11}}{\hat{P}_{11}+\hat{P}_{10}}$$

where $\hat{P}_{11}=\Pr(I_t=1;\, I_{t+1}=1)$ and $\hat{P}_{10}=\Pr(I_t=1;\, I_{t+1}=0)$ are jointed probabilities. Thus, $\hat{p}_{11}$ is a function of $\hat{P}_{11}$ and $\hat{P}_{10}$ and we can write:

$$\hat{p}_{11}= f(\hat{P}_{11},\hat{P}_{10})$$

It is known that, for a given function g, the first-order Taylor series expansion of g(x_0) around x is:

$$g(x_0)=g(x)+g'(x)(x_0-x)$$
$$\Longrightarrow g(x_0)-g(x)=g'(x)(x_0-x)$$

Then, we can write:

$$\begin{aligned}
\widehat{p}_{11}-p_{11} &= \left.\frac{\partial f}{\partial \widehat{P}_{11}}\right|_{(P_{11},P_{10})} . \left(\widehat{P}_{11}-P_{11}\right)+\left.\frac{\partial f}{\partial \widehat{P}_{10}}\right|_{(P_{11},P_{10})} . \left(\widehat{P}_{10}-P_{10}\right)\\
&= \frac{(P_{11}+P_{10})-P_{11}}{(P_{11}+P_{10})^2}\left(\widehat{P}_{11}-P_{11}\right)+\frac{(-P_{11})}{(P_{11}+P_{10})^2}\left(\widehat{P}_{10}-P_{10}\right)\\
&= \frac{P_{10}}{(P_{11}+P_{10})^2}\left(\widehat{P}_{11}-P_{11}\right)-\frac{P_{11}}{(P_{11}+P_{10})^2}\left(\widehat{P}_{10}-P_{10}\right)
\end{aligned}$$

where P_{11} and P_{10} are the asymptotic joined probabilities.

$$\Longrightarrow\sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)=\sqrt{n}\frac{P_{10}}{(P_{11}+P_{10})^2}\left(\widehat{P}_{11}-P_{11}\right)-\sqrt{n}\frac{P_{11}}{(P_{11}+P_{10})^2}\left(\widehat{P}_{10}-P_{10}\right)$$

Asymptotically we have[14]:

$$P_{11}=P_{10}=\frac{1}{4}$$

$$\Longrightarrow \sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)=\sqrt{n}\left(\widehat{P}_{11}-P_{11}\right)-\sqrt{n}\left(\widehat{P}_{10}-P_{10}\right)$$

Then:

[14] For $n\longrightarrow\infty$, we can assume independence and the probabilities become

$$P_{11}=P_{10}=1/4$$

$$Var\left[\sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)\right]=Var\left[\sqrt{n}\left(\widehat{P}_{11}\right)\right]+Var\left[\sqrt{n}\left(\widehat{P}_{10}\right)\right]-2Cov\left[\sqrt{n}\left(\widehat{P}_{11}\right),\sqrt{n}\left(\widehat{P}_{10}\right)\right]$$

▶ **Estimation of** $Var\left[\sqrt{n}\left(\widehat{P}_{11}\right)\right]$ **when** $n\longrightarrow\infty$

$$Var\left(\widehat{P}_{11}\right)=E\left(\widehat{P}_{11}^{2}\right)-E\left(\widehat{P}_{11}\right)^{2}$$

$$E\left(\widehat{P}_{11}^{2}\right)=E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(I_i=1,\ I_{i+1}=1\right).\left(I_j=1,\ I_{j+1}=1\right)\right]$$

The different cases are:

$$j\quad=i;\quad=E\quad\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i=1,\ I_{i+1}=1).(I_i=1,\ I_{i+1}=1)\right]=\frac{n}{n^2}\frac{1}{4}$$

$$j\quad=i-1;\quad+E\quad\left[\frac{1}{n^2}\sum_{i=2}^{n}\left(I_i=1,\ I_{i+1}=1\right).\left(I_{i-1}=1,\ I_i=1\right)\right]=\frac{(n-1)}{n^2}\frac{1}{8}$$

$$j\quad=i+1\quad;\quad+E\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(I_i=1,\ I_{i+1}=1\right).\left(I_{i+1}=1,\ I_{i+2}=1\right)\right]=\frac{(n-1)}{n^2}\frac{1}{8}$$

$$|j-1|>1;\quad+E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{|j-i|>1}^{n}\left(I_i=1,\ I_{i+1}=1\right).\left(I_j=1,\ I_{j+1}=1\right)\right]=\frac{2\sum_{j=2}^{n}(n-j)}{n^2}\frac{1}{16}$$

$$\Longrightarrow E\left(\widehat{P}_{11}^{2}\right)=\frac{n}{4n^2}+\frac{2(n-1)}{8n^2}+\frac{2\sum_{j=2}^{n}(n-j)}{16n^2}$$

$$=\frac{n}{4n^2}+\frac{2(n-1)}{8n^2}+\frac{2(n-1)n-2[\frac{n(n+1)}{2}-1]}{16n^2}$$

$$=\frac{n^2+5n-2}{16n^2}$$

Then:

$$\begin{aligned}Var\left(\widehat{P}_{11}\right)&=E\left(\widehat{P}_{11}^{2}\right)-E\left(\widehat{P}_{11}\right)^{2}\\&=\frac{n^2+5n-2}{16n^2}-\frac{1}{16}\\&=\frac{5n-2}{16n^2}\end{aligned}$$

$$\Rightarrow nVar\left(\widehat{P}_{11}\right) = Var\left(\sqrt{n}\widehat{P}_{11}\right)$$

$$= n\frac{5n-2}{16n^2}$$

$$= \frac{5}{16} - \frac{2}{16n}$$

$$\Rightarrow when \;\; n \longrightarrow \infty, \; Var\left(\sqrt{n}\widehat{P}_{11}\right) \longrightarrow \frac{5}{16}$$

► **Estimation of** $Var\left[\sqrt{n}\left(\widehat{P}_{10}\right)\right]$ **when** $n \longrightarrow \infty$

$$Var\left(\widehat{P}_{10}\right) = E\left(\widehat{P}_{10}^2\right) - E\left(\widehat{P}_{10}\right)^2$$

$$E\left(\widehat{P}_{10}^2\right) = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(I_i = 1, \; I_{i+1} = 0) \, . \, (I_j = 1, \; I_{j+1} = 0)\right]$$

The different cases are:

$$j = i \qquad ; \; = E\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i = 1, \; I_{i+1} = 0) \, . \, (I_i = 1, \; I_{i+1} = 0)\right] = \frac{n}{n^2}\frac{1}{4}$$

$$j = i-1 \qquad ; \; +E\left[\frac{1}{n^2}\sum_{i=2}^{n}(I_i = 1, \; I_{i+1} = 0) \, . \, (I_{i-1} = 1, \; I_i = 0)\right] = 0$$

$$j = i+1 \qquad ; \; +E\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i = 1, \; I_{i+1} = 0) \, . \, (I_{i+1} = 1, \; I_{i+2} = 0)\right] = 0$$

$$|j-1| > 1; \;\; +E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{|j-i|>1}^{n}(I_i = 1, \; I_{i+1} = 0) \, . \, (I_j = 1, \; I_{j+1} = 0)\right] = \frac{2\sum_{j=2}^{n}(n-j)}{n^2}\frac{1}{16}$$

$$\Longrightarrow \; E\left(\widehat{P}_{10}^2\right) = \frac{n}{4n^2} + \frac{2\sum_{j=2}^{n}(n-j)}{16n^2}$$

$$= \frac{n^2+n+2}{16n^2}$$

Then:

$$Var\left(\widehat{P}_{10}\right) = E\left(\widehat{P}_{10}^2\right) - E\left(\widehat{P}_{10}\right)^2$$

$$= \frac{n^2+n+2}{16n^2} - \frac{1}{16}$$

$$= \frac{n+2}{16n^2}$$

$$\Rightarrow nVar\left(\widehat{P}_{10}\right)=Var\left(\sqrt{n}\widehat{P}_{10}\right)$$

$$\begin{aligned} &= n\frac{(n+2)}{16n^2} \\ &= \frac{1}{16}+\frac{2}{16n} \end{aligned}$$

$$\Rightarrow When \;\; n \longrightarrow \infty,\; Var\left(\sqrt{n}\widehat{P}_{10}\right) \longrightarrow \frac{1}{16}$$

▶ **Estimation of** $Cov\left(\sqrt{n}\widehat{P}_{11,}\sqrt{n}\widehat{P}_{10}\right)$ **when** $n \longrightarrow \infty$

$$Cov\left(\widehat{P}_{11,}\widehat{P}_{10}\right)=E\left(\widehat{P}_{11}.\widehat{P}_{10}\right)-E\left(\widehat{P}_{11}\right)E\left(\widehat{P}_{10}\right)$$

$$E\left(\widehat{P}_{11}.\widehat{P}_{10}\right)=E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(I_i=1,\; I_{i+1}=1\right).I\left(I_j=1,\; I_{j+1}=0\right)\right]$$

$$\begin{aligned} j{=}\,i \qquad &;\; =E\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(I_i=1,\; I_{i+1}=1\right).\left(I_i=1,\; I_{i+1}=0\right)\right]=\; 0 \\ j{=}\,i-1 \qquad &;\; =E\left[\frac{1}{n^2}\sum_{i=2}^{n}\left(I_i=1,\; I_{i+1}=1\right).\left(I_{i-1}=1,\; I_i=0\right)\right]\; =0 \\ j{=}\,i+1 \qquad &;\; =E\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(I_i=1,\; I_{i+1}=1\right).\left(I_{i+1}=1,\; I_{i+2}=0\right)\right]\; =\frac{(n-1)}{n^2}\frac{1}{8} \end{aligned}$$

$$|j-1|>1;\;\; =E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{|j-i|>1}^{n}\left(I_i=1,\; I_{i+1}=1\right).\left(I_j=1,\; I_{j+1}=0\right)\right]=\frac{2\sum\limits_{j=2}^{n}(n-j)}{n^2}\frac{1}{16}$$

$$\begin{aligned} \Longrightarrow E\left(\widehat{P}_{11}.\widehat{P}_{10}\right) &= \frac{(n-1)}{8n^2}\;+\frac{2\sum\limits_{j=2}^{n}(n-j)}{16n^2} \\ &= \frac{n-1}{16n} \end{aligned}$$

we have:

$$\begin{aligned} Cov\left(\widehat{P}_{11,}\widehat{P}_{10}\right) &= E\left(\widehat{P}_{11}.\widehat{P}_{10}\right)-E\left(\widehat{P}_{11}\right)E\left(\widehat{P}_{10}\right) \\ &= \frac{n-1}{16n}-\frac{1}{16} \\ &= -\frac{1}{16n} \end{aligned}$$

$$\Rightarrow nCov\left(\widehat{P}_{11,}\widehat{P}_{10}\right) = Cov\left(\sqrt{n}\widehat{P}_{11,}\sqrt{n}\widehat{P}_{10}\right)$$

$$\begin{aligned} &= n.(-\frac{1}{16n}) \\ &= -\frac{1}{16} \end{aligned}$$

$$\Rightarrow When \;\; n \longrightarrow \infty, \;\; Cov\left(\sqrt{n}\widehat{P}_{11,}\sqrt{n}\widehat{P}_{10}\right) \longrightarrow -\frac{1}{16}$$

Then:

$$\begin{aligned} Var\left[\sqrt{n}\left(\widehat{p}_{11}-p_{11}\right)\right] &= Var\left[\sqrt{n}\left(\widehat{P}_{11}\right)\right] + Var\left[\sqrt{n}\left(\widehat{P}_{10}\right)\right] - 2Cov\left[\sqrt{n}\left(\widehat{P}_{11}\right), \sqrt{n}\left(\widehat{P}_{10}\right)\right] \\ &= \frac{5}{16} + \frac{1}{16} - 2(-\frac{1}{16}) \\ &= \frac{1}{2} \end{aligned}$$

B Estimation of $\sqrt{n}\widehat{p}_{00} - p_{00}$ when $n \longrightarrow \infty$

The same developments as before give:

$$Var\left[\sqrt{n}\left(\widehat{p}_{00}-p_{00}\right)\right] = Var\left[\sqrt{n}\left(\widehat{P}_{00}\right)\right] + Var\left[\sqrt{n}\left(\widehat{P}_{01}\right)\right] - 2Cov\left[\sqrt{n}\left(\widehat{P}_{00}\right), \sqrt{n}\left(\widehat{P}_{01}\right)\right]$$

▶ **Estimation of** $Var\left[\sqrt{n}\left(\widehat{P}_{00}\right)\right]$ **when** $n \longrightarrow \infty$

$$Var\left(\widehat{P}_{00}\right) = E\left(\widehat{P}_{00}^2\right) - E\left(\widehat{P}_{00}\right)^2$$

$$E\left(\widehat{P}_{00}^2\right) = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(I_i=0,\; I_{i+1}=0\right).\left(I_j=0,\; I_{j+1}=0\right)\right]$$

$$\begin{aligned} j &= i; \;\; = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(I_i=0,\; I_{i+1}=0\right).\left(I_i=0,\; I_{i+1}=0\right)\right] = \frac{n}{n^2}\frac{1}{4} \\ j &= i-1; \;\; +E\left[\frac{1}{n^2}\sum_{i=2}^{n}\left(I_i=0,\; I_{i+1}=0\right).\left(I_{i-1}=0,\; I_i=0\right)\right] = \frac{(n-1)}{n^2}\frac{1}{8} \\ j &= i+1 \;\; ; \;\; +E\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(I_i=0,\; I_{i+1}=0\right).\left(I_{i+1}=0,\; I_{i+2}=0\right)\right] = \frac{(n-1)}{n^2}\frac{1}{8} \end{aligned}$$

$$|j-1| > 1; \;\; +E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{|j-i|>1}^{n}\left(I_i=1,\; I_{i+1}=0\right).\left(I_j=1,\; I_{j+1}=0\right)\right] = \frac{2\sum_{j=2}^{n}(n-j)}{n^2}\frac{1}{16}$$

$$\Rightarrow E\left(\widehat{P}_{00}^2\right) = \frac{n}{4n^2} + \frac{2(n-1)}{8n^2} + \frac{2\sum_{j=2}^{n}(n-j)}{16n^2}$$
$$= \frac{n^2+5n-2}{16n^2}$$

We have

$$\begin{aligned} Var\left(\widehat{P}_{00}\right) &= E\left(\widehat{P}_{00}^2\right) - E\left(\widehat{P}_{00}\right)^2 \\ &= \frac{n^2+5n-2}{16n^2} - \frac{1}{16} \\ &= \frac{5n-2}{16n^2} \end{aligned}$$

$$\Rightarrow nVar\left(\widehat{P}_{00}\right) = Var\left(\sqrt{n}\widehat{P}_{00}\right)$$

$$\begin{aligned} &= n\frac{5n-2}{16n^2} \\ &= \frac{5}{16} - \frac{2}{16n} \end{aligned}$$

$$\Rightarrow When \;\; n \longrightarrow \infty, \; Var\left(\sqrt{n}\widehat{P}_{00}\right) \longrightarrow \frac{5}{16}$$

▶ **Estimation of** $Var\left[\sqrt{n}\left(\widehat{P}_{01}\right)\right]$ **when** $n \longrightarrow \infty$

$$Var\left(\widehat{P}_{01}\right) = E\left(\widehat{P}_{01}^2\right) - E\left(\widehat{P}_{01}\right)^2$$

$$E\left(\widehat{P}_{01}^2\right) = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(I_i=0,\; X_{i+1}=1).(I_j=0,\; I_{j+1}=1)\right]$$

$$\begin{aligned} j &= i\;; \quad +E\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i=0,\; I_{i+1}=1).(I_i=0,\; I_{i+1}=1)\right] = \frac{n}{n^2}\frac{1}{4} \\ j &= i-1\;; \quad +E\left[\frac{1}{n^2}\sum_{i=2}^{n}(I_i=0,\; X_{i+1}=1).(I_{i-1}=0,\; I_i=1)\right] = 0 \\ j &= i+1\;; \quad +E\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i=0,\; I_{i+1}=1).(I_{i+1}=0,\; I_{i+2}=1)\right] = 0 \end{aligned}$$

$$|j-1|>1; \quad = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{|j-i|>1}^{n}(I_i=0,\ I_{i+1}=1).(I_j=0,\ I_{j+1}=1)\right] = \frac{2\sum_{j=2}^{n}(n-j)}{n^2}\frac{1}{16}$$

$$\Rightarrow E\left(\widehat{P}_{01}^2\right) = \frac{n}{4n^2} + \frac{2\sum_{j=2}^{n}(n-j)}{16n^2}$$

$$= \frac{n^2+n+2}{16n^2}$$

We have:

$$\begin{aligned} Var\left(\widehat{P}_{01}\right) &= E\left(\widehat{P}_{01}^2\right) - E\left(\widehat{P}_{01}\right)^2 \\ &= \frac{n^2+n+2}{16n^2} - \frac{1}{16} \\ &= \frac{n+2}{16n^2} \end{aligned}$$

$$\Rightarrow nVar\left(\widehat{P}_{01}\right) = Var\left(\sqrt{n}\widehat{P}_{01}\right)$$

$$\begin{aligned} &= n\frac{(n+2)}{16n^2} \\ &= \frac{1}{16} + \frac{2}{16n} \end{aligned}$$

$$\Rightarrow When \ \ n \longrightarrow \infty, Var\left(\sqrt{n}\widehat{P}_{01}\right) \longrightarrow \frac{1}{16}$$

▶ **Estimation of** $Cov\left(\sqrt{n}\widehat{P}_{00,}\sqrt{n}\widehat{P}_{01}\right)$ **when** $n \longrightarrow \infty$

$$Cov\left(\widehat{P}_{00,}\widehat{P}_{01}\right) = E\left(\widehat{P}_{00}.\widehat{P}_{01}\right) - E\left(\widehat{P}_{00}\right)E\left(\widehat{P}_{01}\right)$$

$$E\left(\widehat{P}_{00}.\widehat{P}_{01}\right) = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(I_i=0,\ I_{i+1}=0).(I_j=0,\ I_{j+1}=1)\right]$$

$$j=i \quad ; \quad = E\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i=0,\ I_{i+1}=0).(I_i=0,\ I_{i+1}=1)\right] = 0$$

$$j=i-1 \quad ; \quad = E\left[\frac{1}{n^2}\sum_{i=2}^{n}(I_i=0,\ I_{i+1}=0).(I_{i-1}=0,\ I_i=1)\right] = 0$$

$$j=i+1 \quad ; \quad = E\left[\frac{1}{n^2}\sum_{i=1}^{n}(I_i=0,\ I_{i+1}=0).(I_{i+1}=0,\ I_{i+2}=1)\right] = \frac{(n-1)}{n^2}\frac{1}{8}$$

$$|j-1|>1; \quad = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{|j-i|>1}^{n}\left(I_i=0,\ I_{i+1}=0\right).\left(I_j=0,\ I_{j+1}=1\right)\right] = \frac{2\sum_{j=2}^{n}(n-j)}{n^2}\frac{1}{16}$$

$$\Rightarrow E\left(\widehat{P}_{00}.\widehat{P}_{01}\right) = \frac{(n-1)}{8n^2} + \frac{2\sum_{j=2}^{n}(n-j)}{16n^2}$$

$$= \frac{n^2-n}{16n^2} = \frac{n-1}{16n}$$

We have

$$\begin{aligned} Cov\left(\widehat{P}_{00,}\widehat{P}_{01}\right) &= E\left(\widehat{P}_{00,}.\widehat{P}_{01}\right) - E\left(\widehat{P}_{00}\right)E\left(\widehat{P}_{01}\right) \\ &= \frac{n-1}{16n} - \frac{1}{16} \\ &= -\frac{1}{16n} \end{aligned}$$

$$\Rightarrow nCov\left(\widehat{P}_{00},\widehat{P}_{01}\right) = Cov\left(\sqrt{n}\widehat{P}_{00},\sqrt{n}\widehat{P}_{01}\right)$$

$$\begin{aligned} &= n.(-\frac{1}{16n}) \\ &= -\frac{1}{16} \end{aligned}$$

$$\Rightarrow When\ \ n \longrightarrow \infty,\ Cov\left(\sqrt{n}\widehat{P}_{00},\sqrt{n}\widehat{P}_{01}\right) \longrightarrow -\frac{1}{16}$$

We saw that:

$$\begin{aligned} Var\left[\sqrt{n}\left(\widehat{p}_{00}-p_{00}\right)\right] &= Var\left[\sqrt{n}\left(\widehat{P}_{00}\right)\right] + Var\left[\sqrt{n}\left(\widehat{P}_{01}\right)\right] - 2Cov\left[\sqrt{n}\left(\widehat{P}_{00}\right),\sqrt{n}\left(\widehat{P}_{01}\right)\right] \\ &= \frac{5}{16} + \frac{1}{16} - 2(-\frac{1}{16}) \\ &= \frac{1}{2} \end{aligned}$$

C Estimation of θ_s

▶ **The first-order serial correlation is significative: k = 1**

If the first-order of serial correlation is statistically significant but not the second-order one, we have 3 parameters to estimate θ_0, θ_1 and σ_η^2 from the following system of equations:

$$\begin{cases} E\left[X_t^2\right] = (\theta_0^2+\theta_1^2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = \theta_0\theta_1\sigma_\eta^2 \\ 1 = \theta_0+\theta_1 \end{cases}$$

By replacing θ_1 in the first two equations by its value $1 - \theta_0$, we get:

$$\begin{cases} E\left[X_t^2\right] = (\theta_0^2 + (1-\theta_0)^2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = (\theta_0 - \theta_0^2)\sigma_\eta^2 \end{cases}$$

This leads us to:

$$\sigma_\eta^2 = E\left[X_t^2\right] + 2.E\left[X_t.X_{t-1}\right]$$

Thus, we can empirically estimate σ_η^2 from the sample equivalent of $E\left[X_t^2\right]$ and $E\left[X_t X_{t-1}\right]$. The second equation implies that:

$$\frac{E\left[X_t.X_{t-1}\right]}{E\left[X_t^2\right] + 2E\left[X_t.X_{t-1}\right]} = \theta_0 - \theta_0^2$$

Let

$$\gamma_1 = \frac{E\left[X_t.X_{t-1}\right]}{E\left[X_t^2\right] + 2E\left[X_t.X_{t-1}\right]}$$

We get:

$$\theta_0^2 - \theta_0 + \gamma_1 = 0$$

This equation has two solutions:

$$\begin{cases} \theta_{0,1} = \frac{1}{2} + \frac{\sqrt{1-4\gamma_1}}{2} \\ \theta_{0,2} = \frac{1}{2} - \frac{\sqrt{1-4\gamma_1}}{2} \end{cases}$$

This implies that a solution exists if and only if $\gamma_1 \leq \frac{1}{4}$.

Given that $\theta_0 \geq \theta_1$, and both sum to 1, θ_0 is higher than $\frac{1}{2}$, then

$$\theta_0 = \frac{1}{2} + \frac{\sqrt{1-4\gamma_1}}{2}$$

and

$$\theta_1 = 1 - \theta_0 = \frac{1}{2} - \frac{\sqrt{1-4\gamma_1}}{2}$$

We see here that θ_0 is positive, but θ_1 could be negative in certain conditions. Indeed, $\theta_1 < 0$ if

$$\frac{1}{2} - \frac{\sqrt{1-4\gamma_1}}{2} < 0$$
$$\Rightarrow \gamma_1 < 0$$

Then γ_1 should be ≥ 0 to ensure that we have positive weights. We saw that:

$$\gamma_1 = \frac{E\left[X_t.X_{t-1}\right]}{\sigma_\eta^2}$$

The sign of γ_1 depends on the numerator. This means that if $E\left[X_t.X_{t-1}\right] < 0$, it implies $\theta_1 < 0$.We have

$$\begin{aligned} E\left[X_t.X_{t-1}\right] &= E\left[X_t\right].E[X_{t-1}]+Cov(X_t, X_{t-1}) \\ &= Cov(X_t, X_{t-1}) \end{aligned}$$

given that X_t are centered returns. Thereby, if $Cov(X_t, X_{t-1}) < 0$ we will have $\theta_1 < 0$. In other words it means that if the serial correlation of order 1 is negative, not all weights will be positive and the unsmoothing will be incongruous because ξ will be greater than 1, and σ_c^2 will be less than σ_o^2.

Overall, to obtain satisfactory solutions, γ_1 should lead in this interval:

$$0 < \gamma_1 \leq \frac{1}{4}$$

The first order of autocorrelation should not be negative, nor should it be too high.

▶ **The first and the second order of serial correlation are significative: k = 2**

If the first and the second order of serial correlation are both statistically significant, we have 4 parameters to estimate θ_0, θ_1, θ_2 and σ_η^2 from the following system of equations:

$$\begin{cases} E\left[X_t^2\right] = (\theta_0^2+\theta_1^2+\theta_2^2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = (\theta_0\theta_1+\theta_1\theta_2)\sigma_\eta^2 \\ E\left[X_t.X_{t-1}\right] = \theta_0\theta_2\sigma_\eta^2 \\ 1 = \theta_0+\theta_1+\theta_2 \end{cases}$$

The development of the equations gives:

$$\sigma_\eta^2 = E\left[X_t^2\right]+2.E\left[X_t.X_{t-1}\right]+2.E\left[X_t.X_{t-2}\right]$$

We can estimate σ_η^2 empirically from the sample equivalent of $E\left[X_t^2\right]$, $E\left[X_t.X_{t-1}\right]$ and $E\left[X_t.X_{t-2}\right]$.

From the second equation we have:

$$\frac{E\left[X_t.X_{t-1}\right]}{\sigma_\eta^2} = \theta_1 - \theta_1^2$$

$$\Rightarrow \qquad \frac{E\left[X_t.X_{t-1}\right]}{E\left[X_t^2\right]+2E\left[X_t.X_{t-1}\right]+2E\left[X_t.X_{t-2}\right]} = \theta_1 - \theta_1^2$$

Let

$$\delta_1 = \frac{E\left[X_t.X_{t-1}\right]}{E\left[X_t^2\right]+2E\left[X_t.X_{t-1}\right]+2E\left[X_t.X_{t-2}\right]}$$

We get :

$$\theta_1^2 - \theta_1 + \delta_1 = 0$$

This equation has two solutions:

$$\begin{cases} \theta_{1,1} &= \frac{1}{2} + \frac{\sqrt{1-4\delta_1}}{2} \\ \theta_{1,2} &= \frac{1}{2} - \frac{\sqrt{1-4\delta_1}}{2} \end{cases}$$

As pointed by GLM in the Proposition 3 of their model:
(i) $\theta_1 < 1/2$;
(ii) $\theta_1 < 1 - 2\theta_2$

From (i), it follows that:

$$\theta_1 = \frac{1}{2} - \frac{\sqrt{1-4\delta_1}}{2}$$

We also see here that to obtain a satisfactory solution:

$$0 \leq \delta_1 \leq \frac{1}{4}$$

From the value of θ_1 we can get θ_0. From the third equation, we have:

$$\frac{E\left[X_t.X_{t-2}\right]}{E\left[X_t^2\right] + 2E\left[X_t.X_{t-1}\right] + 2E\left[X_t.X_{t-2}\right]} = \theta_0 - \theta_0^2 - \theta_0\theta_1$$

Let

$$\delta_2 = \frac{E\left[X_t.X_{t-2}\right]}{E\left[X_t^2\right] + 2E\left[X_t.X_{t-1}\right] + 2E\left[X_t.X_{t-2}\right]}$$

We get:

$$\theta_0^2 - (1-\theta_1)\theta_0 + \delta_2 = 0$$

This equation has two solutions:

$$\begin{cases} \theta_{0,1} &= \frac{(1-\theta_1)}{2} + \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2} \\ \theta_{0,2} &= \frac{(1-\theta_1)}{2} - \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2} \end{cases}$$

From (ii), we have:

$$\theta_1 < 1 - 2(1 - \theta_0 - \theta_1)$$

$$\Rightarrow \theta_0 > \frac{1-\theta_1}{2}$$

Thus the solution for θ_0 is:

$$\theta_0 = \frac{(1-\theta_1)}{2} + \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2}$$

We also see here that we have a solution if and only if $\delta_2 \leq \frac{(1-\theta_1)^2}{4}$.
Next, we obtain $\theta_2 = 1 - \theta_0 - \theta_1$.This give us

$$\theta_2 = \frac{(1-\theta_1)}{2} - \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2}$$

We can see that θ_2 could be negative in certain conditions. Indeed $\theta_2 < 0$ if

$$\frac{(1-\theta_1)}{2} - \frac{\sqrt{(1-\theta_1)^2 - 4\delta_2}}{2} < 0$$

$$\Rightarrow \delta_2 < 0$$

Thus δ_2 should be ≥ 0 to ensure that we have positive value of θ_2. We saw that:

$$\delta_2 = \frac{E\left[X_t.X_{t-2}\right]}{\sigma_\eta^2}$$

This means that if $Cov(X_t, X_{t-2}) < 0$, in other words if the second order of serial correlation is negative, we will have a negative value for θ_2.

Overall, to obtain satisfactory solutions, δ_1 and δ_2 should lead in these intervals:

$$\begin{aligned} 0 &\leq \delta_1 \leq \frac{1}{4} \\ 0 &\leq \delta_2 \leq \frac{(1-\theta_1)^2}{4} \end{aligned}$$